

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A2

(51) International Patent Classification⁷: **C07K 14/00**

[DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE).

(21) International Application Number: PCT/IB00/01496

(74) Agent: CARPMAELS & RANSFORD; 43 Bloomsbury Square, London WC1A 2RA (GB).

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): GERMAN HUMAN GENOME PROJECT [DE/DE]; Fraunhofer Patentstelle, Leonrodstrasse 68, 80636 Munich (DE).

Published:

— Without international search report and to be republished upon receipt of that report.

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): WIEMANN, Stefan

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/12659 A2

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.

HUMAN DNA SEQUENCES

Background of the Invention

Current methods for testing pharmacological substances rely on a three-stage testing approach to drug development. First, candidate compounds are typically screened in some sort of *in vitro* system, like inhibition of cancer cell growth. Candidates are then tested in an animal model, as a first approximation of systemic effects, including efficacy and toxicity. Compounds that still show promise after these initial *in vivo* screens, finally are tested in humans. Again, human testing typically occurs in three phases: toxicity; preliminary efficacy; and efficacy. The entire process can take more than a decade and cost hundreds of millions of dollars. Aside from the monetary costs and protracted time scale, moreover, current testing regimes waste the lives of countless laboratory animals and needlessly endanger the lives of human subjects.

A need exists, therefore, for more sophisticated drug screening techniques that can be done rapidly *in vitro*. These screening techniques ideally will be reflective of systemic and/or organ-specific responses, so that they provide a reliable indicator of action in a human body. Current techniques, however, tend to utilize only a single or limited number of markers, thus answering only very simple questions that are of questionable medical import. For example, a typical *in vitro* assay may ask whether a lead compound binds a particular receptor, which has been implicated in a certain disorder. It is presumed that such binding is indicative of therapeutic usefulness, but it does not even purport to address systemic effects.

Not only are screening techniques for efficacy inadequate, the available toxicity screens likewise are inadequate. Toxicity, on a first level, is usually measured by animal testing. Aside from the complications related to *in vivo* versus *in vitro* testing, such screens are insufficient because of differences in metabolism, uptake, etc., relative to humans. Thus, improved methods would be not only be *in vitro*-based, they would also be more "human."

With the increasing miniaturization of screening assays and the growing availability of targets for pharmaceutical intervention, there is increasing interest in developing arrays containing large numbers of these targets that can be assayed simultaneously. If such an

array contains a large enough population of targets, it can be used to essentially mimic the systemic response. In other words, the array becomes an *in vitro* surrogate for the human body. The more refined the array, the more accurate the predictive capability. In theory, an array could be constructed that can detect all of the known human expression products simultaneously, thereby, providing a very reliable indicator of the human response to a given compound. These arrays offer advantages over the present *in vitro* screening systems in that they can assay large numbers of responses simultaneously. They are superior to animal testing because they are more "human" and, thus, more predictive of human responses.

In order to construct such arrays, however, the field is in need of further human targets. Advantageously, such targets will be provided with additional physiologically relevant information, such as whether the target is expressed in a particular tissue and whether it is related to a known functional class of targets. In this way, the artisan can focus as needed, for example, on tissue-specific effects or target class-specific effects, thereby providing information useful in evaluating efficacy and/or toxicity.

In addition to a need for pharmacological screening targets, there is a need for further pharmacological substances. These substances can be used in the formulation of medicinal compositions and in treating a wide variety of disorders.

The present invention responds to the aforementioned and other needs in the field by providing a population of novel targets useful, *inter alia*, in the profiling and medicinal contexts described above.

Summary of the Invention

It is an object of the invention, therefore, to provide a set of human cDNA clones. Further to this object, the invention provides sequences of human cDNA clones that were isolated from libraries generated from different human tissues.

It is another object of the invention to provide assemblages of targets useful in profiling matrices for screening pharmacological test compounds. According to this object, assemblages comprising different populations of human nucleic acids, proteins and antibodies are provided. In different embodiments, cDNA library-specific assemblages and target-family-specific targets are provided.

It is a further object of the invention to provide a database of human nucleotide and protein sequences. Further to this object, novel human nucleotide and protein sequences are provided in electronic form. In one embodiment, one or more of these sequences is provided in a searchable database.

It is still another object of the invention to provide biologically active target molecules useful in treating or detecting human disorders. Further to this object, the invention provides nucleic acid and protein molecules that have the capacity to affect disease etiology or symptoms or correlate with known disease states. Also further to this object, a database is provided which comprises the disclosed molecules in electronic form.

It is still a further object of the invention to provide polypeptides encoded by the human cDNA clones disclosed herein. Further to this object, the invention provides antibodies and fragments thereof that are capable of binding to a specific portion of these polypeptides.

It is yet another object of the invention to provide pharmaceutical compositions which comprise an effective amount of a pharmaceutical agent, wherein the pharmaceutical agent is selected from the group consisting of one or more polypeptides contemplated by the invention, variants or functional derivatives thereof, and antibodies thereto; and a physiologically acceptable carrier or excipient.

It is still another object of the invention to provide expression vectors comprising one or more human cDNA clones disclosed herein or fragments thereof; and optionally a promoter operably linked to the cDNA clone or fragment thereof. Further to this object, the invention provides methodology for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by a human cDNA clone disclosed herein.

Detailed Description

The invention results from a need in the art for new human nucleic acids and proteins. This need arises in several contexts. First, there is a need to identify targets for therapeutic intervention. Second, there is a need to identify molecules that may be adversely affected in a therapeutic context, thereby resulting in toxicity. Knowledge of these molecules will aid in

the design of new medicaments with enhanced efficacy and decreased toxicity. Finally, the need encompasses human nucleic acids and proteins that have medicinal applicability in their own right.

In view of these needs, the present inventors set out to isolate and sequence human cDNAs from tissue-specific libraries. In this way, they represent subsets of molecules likely to be targets for therapeutic intervention or for avoiding toxicity. In addition, the inventors divided the molecules into various sub-categories, based on suspected functionality, structural similarity etc, which are of interest from a pharmacological perspective. These molecules are disclosed in provisional application serial nos. 60/149,499 and 60/156,503, filed August 18, 1999, and September 28, 1999, respectively, both of which are hereby incorporated by reference in their entirety.

GENERAL DESCRIPTION OF THE INVENTIVE MOLECULES

The present invention provides novel polynucleotide molecules that, in some instances, have similarities with known molecules. The inventive DNAs were cloned from five different human cDNA libraries. In addition to these DNA molecules, the invention provides their protein translations and antibodies derived from them. The inventive DNA and protein sequences are show individually, below. The inventive nucleic acids also include the complements of these DNA sequences, as well as their RNA counterparts. Methods of producing the molecules also are provided. Further, the invention provides methods for detecting all or part of the molecules and of detecting polynucleotides encoding all or part of the molecules.

The inventive molecules derive from five cDNA libraries: human fetal brain; human fetal kidney; human mammary carcinoma; human testis; and human uterus. For convenience, each sequence bears a designation that indicates from which library it is derived. In particular, these designations are: "hfpbr" for human fetal brain; "hfkf" for human fetal kidney; "hmcfc" for human mammary carcinoma; "htes" for human testis; and "hute" for human uterus. The individual libraries were constructed and screened as described below in the examples.

The protein and DNA molecules of the invention are variously described herein as "target" molecules or "inventive" molecules. The sequences and other information pertinent to the nucleic acid and protein molecules of the invention are shown, below.

Interpreting the data disclosed with the Table and cDNA sequences, below:

The table and data below provide the coding sequences of the inventive cDNAs as well as the protein sequences and other useful information, as set out below.

Grouping

The clones were assigned to the following fourteen functional and/or tissue-derived groups:

1. Cell Cycle
2. Cell Structure and Motility
3. Differentiation/Development
4. Intracellular Transport and Trafficking
5. Metabolism
6. Nucleic Acid Management
7. Signal Transduction
8. Transmembrane Protein
9. Transcription Factors
10. Brain derived
11. Kidney derived
12. Mammary Carcinoma derived
13. Testes derived
14. Uterus derived

Description of Clone Files

The individual clone files are structured in the same pattern. The Sections are separated by paragraphs.

1. Clone Name

The clone names are deciphered with reference to the following example:

DKFZphfkd2_24e23, wherein the code represents:

- producer of library ("DKFZ") (for convenience, this reference may be eliminated)
- a "p" for "plasmid cDNA library" (for convenience, this reference may be eliminated)
- library name (e.g. hfbr = human fetal brain; hfkd = human fetal kidney; hmcfc = human mammary carcinoma; htes = human testes; hute = human uterus)
- an underscore (" _ ") to separate library information from plate information
- plate number (e.g. "16")
- plate coordinates (letter first; e.g. "f14")

2. Group

3. Introduction

short review of the similarities, function of the protein and possible applications

4. Short Information

specifications about the cDNA (who sequenced, completeness of the cDNA, similarity, who sequenced, chromosomal localisation, length of cDNA, localisation of poly A tail and polyadenylation signal)

5. cDNA-Sequence**6. BLASTn Results**

search results of blasting the cDNA sequence against all public databases

7. Medline Entries

information about genes/proteins similar to the novel cDNA (if available)

8. Putative Encoded Protein Information

specifications about the encoded protein (ORF: length and localisation of the reading frame)

9. Protein Sequence**10. BLASTp Results**

search results of blasting the protein sequence against all public databases

11. Pedant Information

output of fully automated annotation: summarises peptide information, homologies, patterns as follows:

[Length]

- length of the protein = number of amino acid residues

[MW]

- molecular weight of the protein

[pI]

- isoelectric point

[HOMOL]

- shows protein with closest similarity to the cDNA-encoded protein

[FUNCAT]

- functional information according to a catalogue developed by Munich

Information center for Protein Sequences (MIPS)

[BLOCKS]

- Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The blocks for the Blocks Database are made automatically by looking for the most highly conserved regions in groups of proteins documented in the Prosite Database. The Prosite pattern for a protein group is not used in any way to make the Blocks Database and the pattern may or may not be contained in one of the blocks representing a group. These blocks are then calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of matches. It is these calibrated blocks that make up the Blocks Database. The WWW versions of the Prosite and SWISS-PROT Databases that are used on this server are located at the ExPASy World Wide Web (WWW) Molecular Biology Server of the Geneva University Hospital and the University of Geneva. World Wide Web URL http://blocks.fhcr.org/blocks/about_blocks.html/ is the entry point to the database.

- here Blocks segments found in the analysed protein sequences are displayed

[SCOP]

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The scop database provides a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in Brookhaven National Laboratory's Protein Data Bank (PDB). It is available as a set of tightly linked hypertext documents which make the large database comprehensible and accessible. In addition, the hypertext pages offer a panoply of representations of proteins, including links to PDB entries, sequences, references, images and interactive display systems. World Wide Web URL <http://scop.mrc-lmb.cam.ac.uk/scop/> is the

entry point to the database. Existing automatic sequence and structure comparison tools cannot identify all structural and evolutionary relationships between proteins. The scop classification of proteins has been constructed manually by visual inspection and comparison of structures, but with the assistance of tools to make the task manageable and help provide generality. Proteins are classified to reflect both structural and evolutionary relatedness. Many levels exist in the hierarchy, but the principal levels are family, superfamily and fold. The exact position of boundaries between these levels are to some degree subjective. Scop evolutionary classification is generally conservative: where any doubt about relatedness exists, we made new divisions at the family and superfamily levels.

- - here SCOPE segments found in the analysed protein sequences are displayed

[EC]

ENZYME is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and it describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided. World Wide Web URL <http://www.expasy.ch/enzyme/> is the entry point to the database.

- here EC-number and name of enzymes with similarity to the analysed protein sequences are displayed

[PIRKW]

- functional information according to the Protein Information Resource (PIR) database catalogue developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[SUPFAM]

- information according to the Protein Information Resource (PIR) database catalogue of protein superfamilies developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[PROSITE]

please refer to 12. PROSITE Motifs

[PFAM]

please refer to 13. PFAM Motifs

[KW]

- overall 2dimensional folding information

- 3D indicates that the proteins is similar to a protein of which a 3 dimensional structure is known

- overall structural information

[]

The last PEDANT-block depicts information about the folding structure of the protein generated by PREDATOR. PREDATOR is a secondary structure prediction program. It takes as input a single protein sequence to be predicted and can optimally use a set of unaligned sequences as additional information to predict the query sequence. The mean prediction accuracy of PREDATOR is 68% for a single sequence and 75% for a set of related sequences. PREDATOR does not use multiple sequence alignment. Instead, it relies on careful pairwise local alignments of the sequences in the set with the query sequence to be predicted. .

World Wide Web URL http://www.embl-heidelberg.de/argos/predator/predator_info.html is the entry point to the database.

- H = helix, E = extended or sheet, _ = coil, T = transmembrane, B = beta

- x indicates a low-complexity region with repeat-like structure which is omitted in all BLAST searches

12. PROSITE Motifs

PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs. World Wide Web URL <http://www.expasy.ch/prosite/> is the entry point to the database. A description of the prosite consensus patterns is also provided, below.

13. PFAM Motifs

PFAM (protein families) is a large collection of multiple sequence alignments and hidden

Markov models covering many common protein domains. World Wide Web URL <http://www.sanger.ac.uk/Pfam/> is the entry point to the database.

Deposit of Clones

Clones were deposited as a pool with the American Type Culture Collection under accession number _____, from which each clone comprising a particular polynucleotide is obtainable. Each clone has been transfected into separate bacterial cells (*E. coli*) in this composite deposit.

The clones may also be obtained from the Resource Center of the German Human Genome Project (Heubner Weg 6, 14059 Berlin, GERMANY). The Resource Center library numbers are slightly different than those presented here, but may be readily obtained by the following key or with the assistance of Resource Center personnel.

The library name becomes a number: brain (hfbr2) becomes 564; kidney (hfkd2) becomes 566; mammary carcinoma (hmcfl) becomes 727; testis (htes3) becomes 434; and uterus (hute1) becomes 586. Next, the plate number is converted to two digits (e.g., "2" becomes "02") and is moved behind the plate coordinate, and the underscore is dropped. The following examples are helpful:

<u>Listed Number</u>	<u>Resource Center Number</u>
DKFZphfbr2_16f21	DKFZp564F2116
DKFZphfkd2_1j9	DKFZp566J091
DKFZphmcf1_1c23	DKFZp727C231
DKFZphtes3_14g5	DKFZp434G0514
DKFZphute1_17k7	DKFZp586K0717

The libraries were constructed using two commercially available vectors. The brain (hfbr2 designations) and kidney (hfkd2 designations) libraries utilize pAMP 1 from Life Technologies and are maintained in XL-2Blue (Stratagene); the uterus (hute1), testes (htes3) and mammary carcinoma (hmcfl) libraries are constructed in pSPORT1, also from Life Technologies, and are maintained in DH10B (Life Technologies). In addition to the following techniques, consultation with the commercial literature available on these clones will make evident all of the housekeeping techniques needed to propagate and isolate the individual constructs. All inserts may be excised with a NotI/SalI digestion. Alternatively, universal primers, flanking the cloning region, may be used to amplify the inserts using PCR methods.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. Methods of probe design are presented below.

Oligonucleotide probes may be labeled with γ - ^{32}P ATP (specific activity 6000 Ci/mmol) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other, non-radioactive labeling techniques can also be used.

Unincorporated label typically is removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe can be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe generally should be approximately 4×10^6 dmp/pmol.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 50 - 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used). The culture should preferably be grown to saturation at 37°C., and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used) and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them. The filter is then preferably incubated at 65°C. for 1 hour with gentle agitation in 6 x SSC (20 x stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 $\mu\text{g/ml}$ of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1×10^6 dpm/mL. The filter is then preferably incubated at 65°C. with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2 x SSC/0.5% SDS at room temperature without agitation, preferably followed by 500 mL of 2 x SSC/0.1% SDS at room

temperature with gentle shaking for 15 minutes. A third wash with 0.1 x SSC/0.5% SDS at 65°C. for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

Alternatively, clones may be grown as described above, and PCR used to isolate the insert DNAs. Methods of PCR are described below and are otherwise well known.

ERROR SCREENING

The DNA sequences found herein derive from individual clones, which are publicly available, as noted above. Thus, the skilled artisan will recognize that any specific sequence disclosed herein readily can be screened for errors by resequencing a particular fragment, in both directions (*i.e.*, by sequencing both strands). Alternatively, error screening can be performed by amplifying and/or cloning any of the inventive DNAs, using for example RT-PCR, and sequencing the resulting amplified product. In the event that there is a sequencing error, reference should be made to the deposited clone as the correct sequence.

USES AND BIOLOGICAL ACTIVITIES OF THE INVENTIVE MOLECULES

The inventive molecules and their derivatives are susceptible to a wide variety of uses, based on functional and/or structural properties. The skilled worker will appreciate, based on the biological activities detailed below, and discussed with regard to the individual sequences disclosed below, that the inventive molecules will find usefulness in numerous therapeutic and diagnostic applications.

The DNA molecules, especially the potassium salts thereof, can be used as fertilizer supplements due to their high nitrogen and phosphorus contents. Since the DNAs are of defined length, they are also useful in gel electrophoresis as molecular weight markers. Due to their similarity with known molecules, certain of the DNA molecules and their variants and derivatives may be used in any number of different diagnostic procedures and therapeutic applications. They may also be used to make the encoded proteins.

The proteins themselves have many possible uses. They may be used as a nutritional supplement for humans, animals and even for laboratory use as, for example, medium for bacterial cultures. Moreover, since the proteins are of defined, known sizes, they may be used as molecular weight markers for gel electrophoresis and gel filtration. Because they are of defined sequences, they also have use in microsequencing and protein fingerprinting applications.

Expression Profiling Applications

Given their known tissue expression and functional associations, assemblages of the inventive proteins (or corresponding antibodies) and nucleic acids are particularly suited to expression profiling applications. Expression profiling generally entails constructing an array of indicators that signal the presence of a particular RNA or protein expression product. Such arrays can be used to evaluate, for example, pharmacological effectiveness and toxicity. In particular, expression profiles from such arrays can be generated from cells treated with known compounds, having known properties, and these profiles can be compared to profiles of unknowns to evaluate similarities and differences, which can be correlated with efficacy or toxicity.

Additional uses of profiling include diagnosis, tracking development, and ascertaining signaling and metabolic pathways. For examples of references describing profiling and its uses, see Farr *et al.*, U.S. Patent 5,811,231 (1998); Seilhamer *et al.*, U.S. Patent 5,840,484 (1998); Rine *et al.*, U.S. Patent No. 5,777,888 (1998); WO 97/27317; WO 99/05323; WO 99/09218; and WO 99/14369. For a device for implementing such techniques, see Lipshutz *et al.*, U.S. Patent No. 5,856,174 (1999) and Anderson *et al.*, U.S. Patent No. 5,922,591 (1999).

In one embodiment, a subset of the inventive DNAs will be arrayed on a substrate, like a gene chip, a filter or a 96-well plate. Test samples containing cells are maintained in the presence of a label capable of incorporation into nascent mRNA. Samples are treated with test and control compounds, which will induce mRNA expression in the sample, resulting in incorporation of label. Whole mRNA is isolated and applied to the array such that it hybridizes with the DNAs contained therein. After washing, the amount of hybridization is quantified and a profile is generated. These steps are repeated with various control and test compounds, thereby generating a library of profiles, which can be used to ascertain the relationships relevant to pharmacological efficacy or toxicity.

The matrices used in such profiling, however, need not be limited to those utilizing DNAs. Rather, other nucleic acids, like RNAs and protein nucleic acids (PNAs), as well as the inventive proteins and antibodies corresponding to the inventive proteins may also be employed. Hence, for example, antibodies could form the array and the samples could be treated in order to label nascent proteins. Whole proteins then would be isolated and applied to the antibody matrix. Developing the resulting signal would result in a protein expression profile, which is useful in essentially the same manner as the nucleic acid profile. A protein matrix could be used, for example, in evaluating antibody responses to pharmaceutical agents in order to eliminate possible cross-reactivity.

Moreover, where nucleic acids are used in the matrix, it is often beneficial to use variants (as defined below) of the molecules described herein. This can be used to account for genetic variations that are of little or no consequence to the function of the resultant gene product. Hence, they can account for wobble or conservative amino acid variations that do not perturb function, like variations in some of the protein motifs elucidated below. Thus, each position in the matrix can employ multiple nucleic acid probes that account for a series of variants.

Expression profiling may also be done, in another embodiment, using two-dimensional protein gels in which the inventive proteins are detected. The resultant profiles can be used in the same way as described.

Matrices useful for profiling may be constructed based on different criteria. Of course, the more relevant profiles will take into account expression of most human genes, preferably all of them. In certain situations, however, it is advantageous to look at a smaller subset. For example, if one were concerned about fetal neural toxicity, a fetal brain-specific matrix might be chosen. On the other hand, if one were interested in targeting mammary carcinoma tissue, a corresponding matrix could be used. Thus, matrices may be constructed using all of the sequences available from a tissue-specific library.

* * *

The following discussion relates to some of the various functional and structural groupings that would be of interest to the artisan wishing to construct profiling matrices. Of course, the artisan will also recognize that these functional descriptions may find additional applicability in the therapeutic and diagnostic applications discussed below.

Cell Cycle

A proliferating cell must coordinate replication and chromosomal separation to ensure that the genome is replicated completely, and that a single copy is correctly inherited by each daughter cell. The cell cycle is the coordinated series of events that achieves these aims. Many of the key events are initiated by a family of conserved Serine/threonine protein kinases, the cyclin-dependent kinases (CDKs), that are activated by the cyclin family of proteins (cyclins A-H). In turn, the cyclin-CDK complexes are modulated by other protein kinases or phosphatases, and by binding specific inhibitor proteins. The enormous variety of ways in which CDK activity can be regulated allows the cell to respond to internal signals generated by preceding events in the cell cycle and to external growth signals.

The somatic cell cycle is divided into four phases: DNA replication (S phase) and chromosome separation (M phase) are separated by gap phases (G1 and G2). At specific control points the decision to begin the next stage (DNA synthesis or mitosis) is carefully regulated.

Cdc2, the primary kinase, is especially required for the G1-S transition and S phase. Cdc4 and Cdc6 are involved at the restriction point, where the cell can decide to proliferate or arrest (G1 \leftrightarrow G0) and Cdc7 is a CDK activating kinase (CAK) as well as a subunit of TFIIH.

The Cyclin-CDK complexes are regulated in various ways. One is through phosphorylation by CDK activating kinases (CAK), like the Y15 kinase (Wee1) and dephosphorylation by CDK associated phosphatases (CAP), like Cdc25A a member of the Cdc25 family (Cdc25A, B and C).

An other way of regulation occurs through two classes of CDK inhibitors (CKI), the INK4 proteins p15, p16, p18, and p19, who negatively regulates the cyclin D CDK complexes and second the p21 family with p21, p27, and p57.

The cell cycle is also regulated through ubiquitin-mediated proteolysis involving the destruction of both cyclins and CDK inhibitors by the 26S proteasome, that requires an ubiquitin conjugating enzyme (UBC) and an ubiquitin ligase. The instability is conferred by PEST regions (cyclin D and E) or a ten amino acid region in the amino terminus (degradation box) in the A- and B-type cyclins.

All these modifications play an important role for the cellular localization, because only the nuclear CDK-cyclin complexes are functional for cell cycle. During G1 phase of the cell cycle, cyclins A, E and D are synthesized and bind to their cyclin-dependent kinase (CDK) partners. CDK complexes containing cyclins A, E and D1 are then imported into and concentrated within nuclei. Cdk6- cyclin D3 has been localized to both cytoplasmic and nuclear compartments, although only the nuclear complex is active. As cells enter S phase, cyclin A and cyclin E complexes remain within the nucleus, whereas cyclin D1 relocates to the cytoplasm for proteolysis at the onset of S phase. Like Cdk2-cyclin A, Cdc2-cyclin A is nuclear and remains so until it is degraded during mitosis. By contrast, as a result of ongoing nuclear import and more rapid re-export, cyclin B1, which binds to Cdc2 upon synthesis during S phase, is predominantly cytoplasmic. Cdc2-cyclin B2 is also cytoplasmic, although this might occur through anchoring of the complex to some cytoplasmic constituent. At prophase, phosphorylation of cyclin B1 promotes accumulation of Cdc2-cyclin B1 in the nucleus, whereas cyclin B2 remains in the cytoplasm until nuclear envelope breakdown.

Two crucial regulators of Cdc2-cyclin B-Wee1 and Cdc25C exist and are responsible for the G2 to M control point. Wee1 is a nuclear protein throughout the cell cycle, whereas Cdc25C binds to 14-3-3 proteins during interphase and remains predominantly cytoplasmic. In some systems Cdc25C, like cyclin B1, rushes precipitously into the nucleus just before entry into mitosis.

The 110-kDa retinoblastoma (tumor suppressor) protein (RB), a pRB-family member is an important regulator of cell-cycle progression and differentiation. Like the E2F family (E2F1-5) or DP family (DP1-3) of transcription activators, RB suppresses inappropriate proliferation by arresting cells in G1 by repressing the transcription of genes required for the transition into S phase. Before the cell proceeds into S phase, RB becomes phosphorylated at multiple sites by the cyclin dependent protein kinases (CDKs) and loses its transcriptional repressing activity. Phosphorylation of RB during late G1 phase results in the dissociation of the E2F-RB repressor complex which allows S-phase specific genes to be transcribed. Cyclin E is the evolutionary conserved target for E2F and interacts together with CDC2 in late G1.

For a proliferating cell it is vital that only undamaged DNA is replicated because if DNA damage is substantial, its replication can lead to chromosome loss or rearrangement.

Thus, we find a G1<->S checkpoint in late G1 that requires tumor suppressor p53. A p53-dependent G1 arrest is effected by the cyclin dependent kinase inhibitor p21 through higher expression levels that inhibits almost all cyclin CDK complexes.

The kinase responsible for phosphorylating the unidentified kinetochore component in metaphase may be a member of the MAP kinase family and appears to be the proto oncogene c-MOS, a cytostatic factor (CSF) in meiosis.

Several categories of proteins are coded for by clones of the invention within the overall group of "Cell cycle" and include, among others, the following:

Tumor suppressors (e.g. N33): Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. The N33 gene has been reported by OMIN OMIN (Online Mendelian Inheritance in Man at <http://www.ncbi.nlm.nih.gov/htbin-post/Omin>) to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) prostate cancer suppression (OMIN *601385). Clones in this category include: fbr2_2k14.

C-TAK1 Cdc25c associated protein kinase: Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. Alterations in the gene coding for the above protein kinase has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Pancreatic cancer (OMIN *60278). Clones in this category include: tes3_7j3.

Cell structure and motility

One of the major differences between prokaryotes and eukaryotes is the ability of the eukaryotic cell to adopt very different shapes dependent on its function during the differentiation process. Animal cells vary from being round to extended cylindric forms like motoneurons or muscle cells. In humans, more than 100 different cell types can be distinguished, each having a characteristic shape. The form of a cell often is closely related to

its capacity to move. Some completely differentiated cells like fibroblasts can still change their form actively, thereby migrating. Other cell types serve as motor elements - "macroscopically" like muscle cells or "microscopically" like ciliated epithelia. Such tasks are fulfilled by a big class of proteins; on the one hand responsible for maintenance of cell structure and contacting neighbor cells or the intercellular matrix and on the other hand for cell motility. These topics cannot be regarded separately: The motility apparatus e.g. must be fixed in the cytoskeleton. Three different types of filaments can be distinguished: Actin filaments, tubulin filaments and intermediate filaments, each present in almost all types of cells.

Actin filaments (F-actin) are built up of monomers (G-Actin). In muscle cells, actin, myosin, for both of which several paralogous genes are known, as well as many more proteins are constituents of the contractile apparatus.

The "thin" and "thick filaments" in a muscle cell consist mainly of actin and myosin, respectively.

Several different proteins are responsible for the anchoring of the actin filaments in the Z-disks (e.g. alpha-actinin and desmin) or at the end of the myofibers in the cell membrane.

Troponin I, -C, -T and Tropomyosin - associated with actin - confer the Ca^{++} -dependent triggering of contraction.

Length of the sarcomere is controlled by the giant protein titin.

In smooth muscle, there is no troponin. Contraction activity is controlled by phosphorylation / dephosphorylation of myosin by a specialized kinase instead. Contractile fibers are not organized in sarcomeres.

Apart from contributing to muscle contraction, the actomyosin system is responsible for many other motions at cellular level, e.g. the amoeboid movement of pseudopodia or the fission of cells at the end of mitosis by a contractile ring.

Besides this, actin fibers fulfill structural tasks like maintenance of the shape of stereocilia or microvilli. Here, actin filaments are connected by proteins like fimbrin. But not

only specialized structures like the mentioned ones contain actin fibers. There is a network covering the complete cell volume with F-actin as a major constituent. Whereas the actin filaments in the structures mentioned above are relatively stable, this F-actin is highly dynamic. Management of the network structure and turnover is achieved by connecting proteins like alpha-actinin, fimbrin or filin; turnover is regulated by gelsolin, villin, and different capping- and fragmentation-proteins.

Microtubules are built up of alpha-beta tubulin heterodimers. Turnover of filaments is achieved by building-in and releasing of monomers with different time constant rates at both ends. The resulting cycle is called "treadmilling". Thirteen strings of tubulin duplets build up one subfiber, whereas one fiber contains two or three of those. A complete axoneme consists of 9 radial and 2 central fibers. This "9+2" - structure is the basis both of flagella, their basal bodies and centrioles. In flagella, several additional structures like radial elements exist. Nexin connects the fibers and dynein is the motor ATPase which shifts the fibers relative to each other. Several genetic diseases like the Kartagener syndrome are caused by deficiencies of distinct proteins in cilia.

Besides this, microtubules are abundant in all types of cells. They are part of a delivery system for organelles, e.g. in the golgi apparatus. A further very important system based on microtubules is the mitotic spindle, it is organized by the centrosomes. Besides many other components, the major part of a centrosome are two centrioles which are built up of nine microtubule-triplets. Most remarkably, new centrioles are not synthesized de novo but generated by duplication of old ones.

Cytoplasmic microtubules are associated with many different proteins. Two major classes are known: The MAPs ("microtubule-associated proteins", with molecular masses between 200 and 300 kD) and the much smaller tau-Proteins with a MW between 60 and 70 kD. These proteins regulate the treadmill-process and the interaction with other structures in the cell.

Besides actin and myosin the so-called intermediate filaments constitute a third class of filaments. In contrast to the former two groups, they do not participate in motility, nor are they dynamic structures subject to a vivid turnover. The most important ones are

neurofilaments (in neurons), keratin filaments (mainly in epithelial cells), and vimentin filaments (in many sorts different cell types).

The biological function of both the cytoskeleton as well as contractile apparatus of a cell does not end at the cell membrane. Cells must be embedded in the extracellular matrix, all cells of a muscle must act as one single mechanical unit and epithelia must resist macroscopic mechanical forces. Hence, cell adhesion and the extracellular matrix are closely connected to the cytoskeleton. Vincullin is one of the proteins which serve as an anchor for intracellular fibers (actin). Different types of desmosomes and tight junctions connect neighbor cells with intercellular fibers. On the inside, cytoplasmic plaques connect them to the cytoskeleton. These structures, on the one hand, serve as mechanical elements whereas gap junctions, on the other hand, connect cells metabolically.

The extracellular matrix consists of a network of proteins, glycoproteins and polysaccharides. Different proteins are present in relation to different mechanical demands: Elastin is found in tissues with high elasticity (lungs, heart) whereas collagen, a more hard-wearing protein, is found in tendons and ligaments. Fibronectin is an extracellular protein highly important for cell adhesion.

Reference: Murray J *et al* (1992): Cell Motil Cytoskeleton 22: 211-223.

Within the overall group of Cell Structure and Motility several categories of proteins are coded for by clones of the invention:

Collagen alpha chain proteins: Proteins with the typical (xxG)_n repeat of collagen proteins and Pfam von Willebrand factor type A domain(s) suggest they are collagen alpha chains. These proteins can find application in modulation of connective tissue, bone and cartilage development and maintenance. OMIN reports collagen alpha chains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Osteogenesis imperfecta, type I (OMIN #166200); 2) Osteogenesis imperfecta congenita (OMIN #166210); 3) Alport Syndrome, X-linked (OMIN #301050); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Ehlers-Danlos Syndrome, Type VII (OMIN #130060); 6) Marfan Syndrome (OMIN #154700); 7) Alport Syndrome, Autosomal Recessive (OMIN #203780); 8) Alpha-2-Deficient Collagen Disease (OMIN 203760); 9) Goodpasture Syndrome (Omin 233450); 10) Osteogenesis Imperfecta,

progressively deforming, with normal sclerae (OMIN #259420); 11)) Ehlers-Danlos Syndrome, Type VII Autosomal Recessive (OMIN *225410); and 12)) Osteogenesis imperfecta, Type IV (OMIN #166220). OMIN reports that von Willebrand factor type A domains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases:: 1) Hemophilia A (OMIN *306700); 2) Von Willebrand Disease (OMIN *193400); 3) Giant Platelet Syndrome (OMIN *231200); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Congenital Thrombotic Diseases due to protein C deficiency (OMIN #176860); 6) Polycystic Kidney Disease 1 (OMIN *601313); 7) Nephrogenic Diabetes Insipidus (OMIN *304800); 8) Factor V Deficiency (OMIN *227400); and 9) Dentatorubral-Pallidoluysian Atrophy (Omin *125370). Clones in this category include: fbr2_2b5.

Radial spokehead protein: Radial spokehead proteins, e.g., Chlamydomonas reinhardtii radial spokehead protein of flagella or axoneme and the Strongylocentrotus purpuratus sea urchin spermatozoa protein p63, and human proteins with similarity thereto are important for the maintenance of a planar form of sperm flagellar beating. The human protein(s) can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men (e.g., in sterility). Clones in this category include: tes3_15i5.

Ankyrins: Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus these proteins are involved in coupling of cyto skeleton and cell membrane. OMIN reports that Ankyrins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Hereditary Spherocytosis (OMIN *182900); 2) Hemolytic Poikilocytic Anemia due to reduced ankyrin binding sites (OMIN 141700); 3) Atypical Elliptocytosis (OMIN 225450); 4) Autosomal recessive spherocytosis (OMIN #270970); 5) Werner Syndrome (OMIN *277700); and 6) Rhesus-unlinked type Elliptocytosis (OMIN #130600). Clones in this category include: tes3_18i7.

FGD1-related F-actin binding protein (Farbin/FGD1): FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. (OMIN 305400). Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as

described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. Clones in this category include: tes3_72k15.

Paramyosins: Paramyosin is a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. Clones in this category include: tes3_7b22.

Tuftelin: Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix. The new protein can find application in modulation of tissue-calcification, especially the uterus. As reported by OMIN, tuftelin has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with amelogenesis imperfecta (OMIN *600087). Clones in this category include: utel_19g22.

Cell Adhesion Regulator (CAR1): CAR1 is involved in the regulation of cell-cell adhesion. OMIN reports the association (as potentially diagnostic, therapeutic, causative, and/or related, etc...) of CAR1 with tumor suppression by the reduction of tumor invasion (OMIN *116935). Clones in this category include: utel_24j6.

Differentiation/Development

Almost every multicellular organism originates from meiotic cell divisions and the recombination of a paternal and a maternal set of chromosomes. After fertilization of the egg, all cells of a body originate from this one cell. Thus the cells of the developing body are initially genetically alike. But phenotypically they become very different. They are specialized to a certain cell type and arranged in an organized pattern to a certain type of tissue and the whole structure has the well-defined shape of an organ. All these features are determined by the DNA sequence of the genome, which is reproduced in every cell. Each cell acts on the genetic instructions given to a certain time and at a certain place of development and plays its individual part in the multicellular organism. Cell differentiation may be divided into three general steps: cell cycle exit, apoptosis protection and tissue specific gene

expression. These processes are coordinated to provide the final and unique tissue characteristics.

An animal cell that has achieved a certain level of development is said to be determined. This differentiation of a cell may be irreversible and in that case the cell may be renewed only by simple duplication. Other cells are renewed by means of stem cells which are immortal (e.g. stem cells of the bone marrow, epidermal stem cells). The genetic control of development is extensively studied in non-vertebrates and vertebrates. The classical animal model is the fruit fly *Drosophila* and the modern model is the transgenic mouse. Animal transgenesis has proven to be useful for physiological as well as physiopathological studies. Besides the approach based on the random integration of a DNA construct in the mouse genome, gene targeting can be achieved using totipotent embryonic stem cells for targeted transgenesis. Transgenic mice are then derived from the embryonic stem cells. This allows the introduction of null mutations in the genome (so-called knock-out) or the control of the transgene expression by the endogenous regulatory sequence of the gene of interest (so-called knock-in). Mice can be created that express wild-type genes, mutant genes, marker genes or cell lethal genes in a tissue specific manner. These animal models allow to follow changes in tissue and organ development and lead to a better understanding of the cellular function of many genes or to the generation of animal models for human diseases. Fundamental problems in immunology, onset and development of cancer, regulation in fatty acid metabolism, aspects of cardiovascular function, control of the central nervous system development, analysis of reproductive development and function are only some examples of research interests.

The final stage of cell differentiation is growth arrest. In animal tissues with rapid cell turnover terminally differentiated cells undergo programmed cell death. The cells have the ability to kill themselves by activating an intrinsic cell suicide program when they are no longer needed or have become seriously damaged. The execution of this program is termed apoptosis. Apoptosis is of importance for development and homeostasis of animals. The key components of this program have been conserved in evolution from worms (*C. elegans*) to insects (*Drosophila*) to humans. The roles of apoptosis include the sculpting of structures during development, deletion of unneeded cells and tissues, regulation of growth and cell number, and the elimination of abnormal and potentially dangerous cells. In this way

apoptosis provides "quality control mechanism" that limits the accumulation of harmful cells, such as virus-infected cells and tumor cells. On the other hand inappropriate apoptosis is associated with a wide variety of diseases, including AIDS, neuro-degenerative disorders and ischemic stroke. Because it is now clear that apoptosis is a result of an active, gene-directed process, it should be eventually possible to manipulate this form of cell death by developing drugs that interact with its recently identified mechanisms of action. Inducers of cell differentiation, cell cycle arrest and apoptosis might be the novel molecular targets for new anticancer agents in addition to the signaling pathways for growth factors and cytokines.

Proteins, factors, receptors and genes of importance in apoptosis:

Proteases:

- Calpain, an intracellular cysteine protease, exact role unknown.
- Caspase-1 to Caspase-11, a family of proteases synthesized as an inactive proenzyme. Targets of the activated enzymes include: poly(ADP-ribose) polymerase, DNA-dependent protein kinase, U1 ribonucleoprotein, nuclear laminins and cytoskeleton components (actin).
- Granzyme B, a serine protease released by cytotoxic T-cells.

Receptors:

- CD 95 (synonyms: Fas, APO-1), a receptor protein of the TNF-receptor family which includes TNF-R1 and TNF-R2 with the common characteristic of a 70 amino acid cytoplasmic domain.
- FADD (synonym: MORT-1), a cytoplasmic protein
- DR-3 (synonym: APO-3) a member of the TNF-receptor-family
- DR-4 and DR-5

Genes:

- ced-3, ced-4 and ced-9 encode the general apoptotic and antiapoptotic program in *Caenorhabditis elegans*. Apaf-3 is the mammalian homologue of ced-3.

- Bcl-2 / Bcl-xL / Bax / Bcl-xS / Bak: a large gene family that can either inhibit or promote apoptosis.

- Cytokine response modifier A, a cowpox virus gene whose gene product inhibits caspases.

Others:

- Caspase-activated DNase (CAD) and its inhibitor (ICAD), causes DNA fragmentation in the nucleus

- Ceramide, a complex lipid that acts as a second messenger.

- c-Jun N-terminal kinase (JNK) is a proline-directed kinase

- p53 protein, is essential for the induction of apoptosis as a response to chromosomal damage.

- RAIDD, a death signal-transducing protein.

- Receptor interacting protein (RIP) is an accessory protein with a death domain and a serine/threonine kinase activity.

- Sphingomyelinase, an enzyme that hydrolyzes the complex lipid sphingomyelin to ceramide.

- Tumor necrosis factor (TNF) is a type -II membrane protein

- TNF-receptor associated factor (TRAF2), is an accessory protein that can bind to both TNF-R1 and TNF-R2.

Within the overall group of Differentiation/Development, several categories of proteins are coded for by clones of the invention:

Interleukins (e.g. Interleukin-7): Interleukin precursors related to interleukin-7, for example, are expected to act as new growth factors for human B lineage cells. Additionally,

these proteins should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. These interleukins could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells. (OMIN *146660). Clones in this category include: tes3_35e21.

Testis-specific Y-encoded proteins: The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. Proteins of the TSPY-SET-NAP1L1 family represent proteins closely related to TSPY. These proteins seem to be involved in early spermatogenesis. Clones in this category include: fbr2_2d15.

Intracellular transport and trafficking

Eukaryotic cells rely for their viability on the partitioning of many basic cellular processes into membrane-bounded organelles. These are the nucleus, endoplasmic reticulum (ER), Golgi apparatus, endosomes, lysosomal compartments, mitochondria and peroxisomes. Most molecules destined for the lysosome, cell surface and outside the cell are routed through the ER and Golgi, which together with the vesicular intermediates between them, comprise the secretory pathway (Palade 1975). In the ER and Golgi compartments proteins are sorted, modified and often assembled into complexes *en route* to their final destination. Incorrectly assembled proteins are retained in the ER until they fold correctly or are targeted for degradation. Additional proteins are translocated into and function within the luminal spaces of organelles or are secreted. Thus a large proportion of proteins synthesized require targeting to membranes either for insertion into or transport across them. A major purpose of this is growth. The secretory pathway is dependent on an intact cytoskeleton and also closely linked to general metabolism by affecting ribosome biogenesis (Mizuta and Warner, 1994). A huge number of proteins is required for targeting, translocation and sorting of newly synthesized proteins.

The first step in sorting is the recognition of cis-acting targeting or signal sequences that organelle-targeted proteins contain. This is carried out by cytosolic targeting factors and/or receptors on the membrane to which the protein is targeted. In some cases the primary

sequences are extremely degenerate, with only the overall character being conserved (hydrophobicity for an ER signal sequence, helical amphiphilicity for mitochondrial targeting sequence (Kaiser *et al.*, 1987; Lemire *et al.*, 1989). Following the targeting step, proteins are either inserted into or transported across the membrane (translocated) through a proteinaceous apparatus (termed the translocon). The translocon include or recruit motors to drive the translocation process in the correct direction (Schatz and Dobberstein, 1996).

Defined intracellular protein transport steps:

- ER
 - targeting to the ER
 - translocation into the lumen of the ER, and, depending on the presence of certain signals in the peptide sequence transport through the golgi complex
- Mitochondria
 - targeting
 - translocation
- Peroxisomes
- The general secretory pathway
 - protein modification, assembly and quality control in the ER
 - vesicle-mediated trafficking
 - vesicle docking and fusion
 - transport through the golgi apparatus and sorting at the trans-golgi
 - transport to the cell surface
 - transport routes to the lysosome
- Endocytosis
- Specialized protein transport routes
- Protein export from the cytoplasm

References: Palade, G (1975) Science 189:347-358; Mizuta et al. (1994) Mol Cell Biol 14: 2493-2502; Kaiser *et al.* (1987) Science 235: 312-317; Lemire *et al.* (1989) J Biol Chem 264: 20206-20215; Schatz et al. (1996) Science 271: 1519-1526.

Rab proteins

In eukaryotic cells the compartmentalisation of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins

and other molecules. Trafficking between organelles within the secretory pathway occurs as vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation

inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle, most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with α -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca^{2+} -binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca^{2+} influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A

homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991) Mol. Cell. Biol. 11, 872-885; Echard et al. (1998). Science. 279, 580-585; Geppert et al. (1998) Annu. Rev. Neurosci. 21, 75-95; Guo et al. (1999). EMBO J. 18, 1071-1080; Kato et al. (1996) J. Biol. Chem. 271, 31775-31778; Novick et al. (1997) Curr. Opin. Cell Biol. 9, 496-504; Peterson (1999) Curr. Biol. 9, 159-162; Poirier et al. (1998) Nat. Struct. Biol. 5, 765-769; Vitale et al. (1998) EMBO J. 17, 1941-1951; Wang et al. (1997) Nature. 388, 593-598; Yang et al. (1999) J. Biol. Chem. 274, 5649-5653.

Within the overall group of Intracellular Transport and Trafficking several categories of proteins are coded for by clones of the invention.

Rab proteins:

Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells. . Clones in this category include: fbr2_2i17, fbr2_3b16.

Rab10 appear concentrated on membranes in the perinuclear region. Rab 10 has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Choroideremia (OMIN *303199); and 2) RETT Syndrome (OMIN 312750). Clones in this category include: fbr2_62l19.

In mice, Rab17 shows epithelial cell specificity. Rab 17 is discussed as candidate gene for the mouse mutations In (leaden), Tw (twirler), and ax (ataxia). Cloned from a brain cDNA library, the new putative Rab-protein is expected to be involved in vesicle trafficking within neuronal cells. These proteins can find application in modulating the transport of vesicles inside neuronal cells, which are essential for development of functional dendritic processes. . . Clones in this category include: fbr2_41m15.

Ankyrin G: The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments. Ankyrin G has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Werner disease (OMIN *277700). Clones in this category include: fkd2_24p5.

Zn-T-transporters: The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide. These proteins can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation. (OMIN *602878, *602095). Clones in this category include: fbr2_62f10.

Metabolism

This group includes proteins which are involved in the uptake and consumption of nutrients, and enzymes which are part of the biochemical pathways for energy metabolism or

which are involved in the supply of building blocks of nucleic acids, proteins (NTPs, dNTPs, amino acids) for DNA/RNA and protein synthesis, and fatty acids (membranes), to allow for the generation of higher order structures. This group constitutes the most important and largest group in prokaryotes and lower eukaryotes. The higher the evolutionary level of an organism is, however, the more other protein classes like 'signal transduction', 'cell cycle' and 'differentiation and development' increase in importance and number of representatives.

Proteins involved in the metabolism of energy and compounds (here: other than nucleic acids or proteins) are usually the products of house keeping genes, they are often constitutively and/or ubiquitously expressed.

Several categories of proteins are coded for by clones of the invention within the overall group of Metabolism:

NAT1, ARD1: In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. These can find application modulating NAT assembly and action and therefore could be important in metabolism of drugs and environmental mutagens.(OMIN *108345). Clones in this category include: fbr2_3g8.

Apolipoprotein E receptor: In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. These proteins can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins. In normal individuals, chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. In familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants because of a defect in apolipoprotein E. Accumulation of the remnants can result in xanthomatosis and premature coronary and/or peripheral vascular disease. OMIN reports that apolipoprotein has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Familial combined hyperlipidemia (OMIN 144250); and 3) Alzheimer disease. (OMIN #104300). Clones in this category include: fbr2_62017.

Ubiquitin carboxyl-terminal hydrolases: Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. OMIN reports that Ubiquitin-specific proteases have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Lung carcinoma (OMIN *603486); 2) x-linked retinal diseases (OMIN *300050); 3) oncogenesis (OMIN *300050); 4) ovarian cancer (OMIN *300050). Clones in this category include: fbr2_78k24; htes3_27d1.

Phosphoserine signature (phosphoglucosyltransferases, phosphomannosyltransferase): These proteins take part in the conversion of hexose phosphates. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Fanconi-Bickel Syndrome (OMIN #227810). Clones in this category include: fkd2_24b15.

NADH ubiquinone oxidoreductase: NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Brancio-oto-renal syndrome (OMIN *6601445). Clones in this category include: fkd2_3o17.

Transketolases: Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: Wernicke-Korsakoff Syndrome (OMIN *277730). Clones in this category include: tes3_17i17.

Fatty acid-CoA synthetases/ligases: These proteins contain AMP-binding domain signature(s), which is present in enzymes which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic,

causative, and/or related, etc...) with the following diseases: 1) Alport syndrome , mental retardation and elliptocytosis (OMIN *300157); 2) Adrenoleukodystrophy (OMIN *300100). Clones in this category include: tes3_35k17.

ADP/ATP or Adenine Nucleotide Translocators: These proteins contain mitochondrial energy transfer signature(s) and are most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) cardiomyopathy (OMIN *103220); 2) myopathy (OMIN *103220); 3) Progressive external ophthalmoplegia (OMIN *601227). Clones in this category include: tes3_35n12.

Carboxylesterases: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hepatic carboxylesterase with detoxification of foreign compounds (OMIN *114835); 2) non-Hodgkin lymphoma (OMIN *114835); 3) B-cell chronic lymphocytic leukemia (OMIN *114835); 4) rheumatoid arthritis (OMIN *114835). Clones in this category include: tes3_35n9.

Heat shock proteins: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) 27 kd heat shock protein has been correlated with thermotolerance in response to environmental challenges and developmental transitions. (OMIN *6021295). Clones in this category include: utell_23e13.

Nucleic acid management

The genetic information is stored in the form of nucleic acids in all organisms. Two kinds of nucleic acids exist, DNA and RNA. Whereas the more stable DNA in most organisms constitutes the storage form of the genetic information, the labile RNA and in particular mRNA is an intermediate used for the temporal expression of specific genes.

In eukaryotes, DNA is usually a double stranded linear molecule consisting of two antiparallel strands and made up of a deoxyribose, a phosphorus backbone and the four bases A, C, G, and T. The DNA of some organisms has a ring structure. The structure of DNA was

unraveled years ago by Watson and Crick. DNA is directional molecule determined by the C-atoms of the sugar.

The most important processes dealing with nucleic acids are:

- replication (e.g. DNA polymerases, Telomerase)
- transcription (RNA polymerases)
- RNA processing (maturation - splicing and degradation)
- in addition, enzymes and proteins exist which require a nucleic acid (mostly RNA) in the active center to be functional (ribozymes - e.g. RNase, Ribosomal proteins)

The DNA of a cell is replicated in the S-phase of the cell cycle. Several enzymes carry out the task of doubling this nucleic acid. As all steps of the cell cycle, also the process of replication is tightly regulated. The enzyme DNA polymerase and several other proteins are involved in this process. Whereas many prokaryotes do have only one origin of replication (i.e., the starting point of the replication cycle), in eukaryotic DNAs (chromosomes) multiple such start points exist. The switch from the synthesis (S) phase to the subsequent G2 or M phases of the cell cycle are dependent on the completion of the replication. This makes clear, that a number of proteins are involved in the replication itself as well as in the control of the process. Since most eukaryotic chromosomes are linear structures, additional proteins and enzymes are necessary to make sure that the structure is maintained through successive generations. This includes those proteins necessary to build the three dimensional structure of chromosomes (e.g. histones) and the structural network of the nucleus and nucleolus (including the defined localization of transcriptionally active genes in the vicinity of nucleoli) but also such enzymes as telomerase which guarantees the integrity of the chromosomal ends.

The expression of genes is usually performed in two steps. First a messenger RNA (mRNA) is produced (transcribed) in one to many copies and second this mRNA is translated into the protein product. The regulation of transcription is discussed under the separate heading 'transcription factors', but also the classes 'signal transduction', 'development', 'cell cycle' and others are affected as the expression of certain genes determines the fate of a cell or organism.

The primary transcript (hnRNA - heterogeneous nuclear RNA) is a single stranded one-to-one copy of the gene as it is located on the chromosome. Before a protein can be translated, already during transcription the process of maturation is initiated. Firstly, a 5' cap structure is enzymatically and covalently added to the RNA, blocking the 5' end of the RNA.

Second, when the RNA polymerase has terminated polymerization, the enzyme poly A polymerase adds varying numbers of adenine residues to the 3' end of the transcript. This enzyme recognizes the sequence AAUAAA or AUUAAA (+ some minor variations), cuts the RNA 10 - 30 nucleotides downstream and adds the A residues. The size of the poly A sequence affects the stability of the RNA. Finally, in the process of splicing, the introns present on the genomic level and also present in the hnRNA are spliced out by a multi-protein complex consisting of several proteins and RNAs. The finally matured mRNA is exported to the cytoplasm where it is translated with help of the ribozymes.

The half life of RNA is usually much shorter than that of DNA. Usually, the mRNA is degraded shortly after synthesis, to guarantee a very defined window of expression of a given gene. This regulation is necessary to specifically maintain or change the set of proteins present at any time in a cell. Specific regions in the 3'UTR (untranslated region) determine the stability of the mRNA in the cytoplasm before it is degraded by RNases, enzymes consisting both of protein and RNA.

References: Watson and Crick (1953) *Nature* 171: 737-738.

Several categories of proteins are coded for by clones of the invention within the overall group of "Nucleic acid management" and include, among others, the following:

RNA helicases including DEAD/H box helicases: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. DEAD box proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by with the following disease processes and/or genes: 1) ataxia-telangiectasia gene: "A human gene (DDX10) encoding a putative DEAD-box RNA helicase at 11q22-q23" *Genomics* 33:199-206, 1996, Savitsky et al., (OMIN *601235); 2) hematopoietic tumors: "Cloning and expression of a murine cDNA homologous to the human RCK/P54, a lymphoma-linked chromosomal breakpoint 11q23", *Gene* 166:293-6, 1995, Seto et al. (OMIN *600326); 3) dermatomyositis: a) "The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation."

Arthritis Rheum. 38: 1389-1399, 1995, Seelig et al. (OMIN *603277); b) "Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen." (Letter), *Arthritis Rheum.* 39: 1769-1771, 1996., Seelig et al. (OMIN *603277); c) "The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities", *Cell* 95: 279-289, 1998. Zhang et al. (OMIN *603277); 4) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200); 5) Mucopolysaccharidosis Type IVA (OMIN *253000); 6) Albinism I (OMIN *203100); 7) Wilms Tumor 1 (OMIN *194070); 8) Spinocerebellar Ataxia 7 (OMIN *164500). Clones in this category include: fbr2_23b10, fbr2_3cl8, fbr2_6ol7, fbr2_82i24, and tes3_14h21.

Inorganic pyrophosphatase: Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Clones in this category include: fbr2_64a15.

DNA-damage-inducible protein (dinP) or Proteins induced by DNA-Damage: The dinB/P pathway is a second SOS-pathway in E.coli. Genes related to this seem to be involved in modulating DNA repair and mutagenesis. Clones in this category include: fbr2_72b18.

Proteins with myc-type, helix-loop-helix dimerization domain signature(s): This helix-loop-helix domain mediates protein dimerization has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, these proteins could be novel DNA-binding proteins. Clones in this category include: fbr2_72l12.

Cytosolic ribosomal proteins L36: L36 seems to be part of the eukaryotic ribosomal peptidyl transferase center and can find application in modulation of ribosome assembly, maintenance and activity. Clones in this category include: fkd2_3b2.

Ribonuclease H: Ribonuclease H proteins are RNA modifying proteins and have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Adenomatous Polyposis of the Colon (OMIN

*175100); 2) Retinoblastoma (OMIN *180200) ; and 3) Von Hippel-Lindau Syndrome (OMIN *193300). Clones in this category include: phtes3_15j3.

Signal transduction

Cells in higher order organisms need to continuously communicate with its environment especially with other cells of the same organism in order to maintain the function and specialization of the whole system these cells are part of. This important task of communication is performed with help of cell-surface receptors which receive and transmit signals from outside into the cell.

G-proteins

The largest known family of cell-surface receptors is that of the G-protein-coupled receptors, which mediate the transmission of diverse stimuli such as neurotransmitters, glycopeptides, hormones, peptides, odorant molecules, and photons. The functional unit of these receptors is composed of the receptor molecule itself (GPCR) which is anchored in the cytoplasmic membrane with seven membrane spanning domains, the heterotrimeric G-protein which is composed of α and $\beta\gamma$ -subunits ($G\alpha$ and $G\beta\gamma$), and the effectors that interact with $G\alpha$ and / or $G\beta\gamma$. In particular, the dissociated $G\alpha$ and $G\beta\gamma$ can regulate the activities of a number of effector molecules such as adenylate cyclases, phospholipase C isoforms, ion channels, and tyrosine kinases, resulting in a variety of cellular functions. The process of signal transduction must be tightly regulated and reversible in order to avoid overstimulation, to achieve signal termination, and render the receptor responsive to subsequent stimuli [Iacovelly L. et al., (1999) *FASEB J.* 13, 1-8, Hamm, H.E. (1998) *J. Biol. Chem.* 273, 669-672].

G-proteins are GTPases that, upon binding of GTP change their conformation which in return unmasks structural motives, in particular the so called effector loop, which can mediate the interactions to target proteins, or effectors, for the GTPases. This ability enables the GTPases to cycle between active, GTP-bound and inactive, GDP bound conformations and in the process to function as molecular traffic lights in a multitude of signal transduction pathways. The most important of these signal transduction pathways that are regulated with help of G-proteins are that of the phospholipase C / protein kinase C and that of the adenylate cyclase / protein kinase A.

The cycling of GTPases is tightly regulated by three main classes of proteins: The exchange of hydrolyzed GDP for a fresh GTP is facilitated by guanosine nucleotide exchange factors (GEFs), the hydrolysis of GTP to GDP is sped up by GTPase-activating proteins (GAPs), and the dissociation of GDP from the GTPases is inhibited by GDP dissociation inhibitors (GDIs) [Tapon and Hall (1997) *Curr. Opin. Cell. Biol.* **9**, 86-92, Van Aelst and D-Souza-Schorey (1997) *Genes Dev.* **11**, 2295-2322].

SOC-family

A conserved motif that was originally identified in proteins that negatively regulate the signaling action of cytokines was termed SOCS box, the Suppressor Of Cytokine Signaling. Based on homology, five distinct structural protein classes have been identified since that carry this motif. The function of most of these proteins is presently not known. Common to the proteins is only the SOCS box which is located near the C-terminus of the respective peptides. Recently, the SOCS box has been demonstrated to induce binding of proteins to elongins B and C which could target the proteins (and bound substrates) to the proteasomal protein degradation pathway (Kamura, T. *et al.* (1998) *Genes Dev.* **12**, 3872-3881; Zhang, J.-G. *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2071-2076).

The class where the SOCS box was originally described contains several members (SOCS-1-SOCS-7 and CIS). In addition to the SOCS box, these proteins also contain a SH2 (Src-homology 2) domain and a variable N-terminus. These SOCS proteins appear to form part of a classical negative feedback loop that regulates cytokine signal transduction. Upon cytokine stimulation, expression of SOCS proteins is rapidly induced and the proteins inhibit further cytokine action. The mode of action of the SOCS proteins is variable. While SOCS-1 binds and inhibits the JAK (Janus kinases) family of cytoplasmic protein kinases [Narazaki M. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13130-13134, Nicholson, S.E. *et al.* (1999) *EMBO J.* **18**, 375-385], CIS appears to act by competing with signaling molecules such as the STATs (Transducers and Activators of Transcription) family for binding to phosphorylated receptor cytoplasmic domains [Yoshimura, A. *et al.* (1995) *EMBO J.* **14**, 2816-2826; Matsumoto, A. *et al.* (1997) *Blood* **89**, 3148-3154].

A second class of SOCS box protein contains additionally WD-40 repeats which were initially identified in the mouse WSB-1 and -2 proteins. The functions of WD-40 proteins are not completely understood but seem to be rather divergent. In Cdc4p the WD-40 repeats probably are necessary for binding the substrate for Cdc34p [Mathias, N. *et al.* (1999) *Mol.*

Cell Biol. **19**, 1759-1767]. Cdc4p is a component of a ubiquitin ligase that tethers the ubiquitin-conjugating enzyme Cdc34p to its substrates. The posttranslational modification of a protein by ubiquitin usually results in rapid degradation of the ubiquitinated protein by the proteasome. The transfer of ubiquitin to substrate is a multistep process where WD-40 repeats might play an important function.

Other WD-40 containing proteins (e.g. the retino blastoma binding protein RbAp48) have been shown to bind metal ions (Zinc) and that this metal binding might mediate and/or regulate protein-protein interactions which are functionally important in chromatin metabolism [Kenzior, A.L. and Folk, W.R. (1998) *FEBS Lett.* **440**, 425-429]. These proteins are involved in the RAS-cAMP pathway that regulates cellular growth [Ach R.A. *et al.* (1997) *Plant Cell* **9**, 1595-1606].

The SPRY domain has been identified in pyrin or marenostin, a protein which is mutated in patients with Mediterranean fever and which is similar to the butyrophilin family. While butyrophilins seem to be involved in the lactation process in mammals, the function pyrin is unknown. Three proteins (SSB-1 to -3) have been identified to contain both SPRY and SOCS box motifs. The function of these proteins is also not known.

Ankyrin repeat containing proteins share a 33-residue repeating motif, an L-shaped structure with protruding β -hairpin tips which mediate specific macromolecular interactions with cytoskeletal, membrane, and regulatory proteins. These proteins play fundamental roles in diverse biological activities including growth and development, intracellular protein trafficking, the establishment and maintenance of cellular polarity, cell adhesion signal transduction, and mRNA transcription. Three proteins that contain ankyrin repeats (ASB-1 to -3) have been identified to contain a C-terminal SOCS box additionally to the ankyrin repeats. The function of these proteins or the individual domains remains to be discovered [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

A few small GTPases (RAR and RAR like) do also contain a SOCS box. GTPases are involved in signal transduction during cellular communication. The function of the SOCS box in this type of proteins is currently unclear [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

Ca²⁺ as second messenger

The bivalent cation Ca²⁺ is, besides cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very

low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

cAMP

The cyclic AMP is produced by the enzyme adenylate cyclase in response to extracellular signals. Certain G-proteins stimulate the activity of adenylate cyclase which converts ATP to cAMP and PPI. Two molecules of cAMP bind to each of two regulatory subunits of cAMP dependent protein kinase which in turn dissociate from the two catalytic subunits of the heterotetramer R_2C_2 . Upon release of the C-subunits, they become active and phosphorylate substrate proteins at Ser and Thr residues. The process leading from binding of extracellular molecules to their receptors, the transmission of the stimuli into the cell, the activation of adenylate cyclase and the subsequent activation of cAMP dependent protein kinase is one of two major signal transduction pathways in eukaryotic cells. Since the phosphorylation of proteins is a posttranslational modification of proteins, the kinases are described in the class "signal transduction."

SARA

Members of the transforming growth factor β (TGF β) superfamily signal through a family of cell-surface transmembrane serine/threonine kinases, known as type I and type II receptors (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Ligand induces formation of heteromeric complexes of these receptors, and signaling is initiated when receptor I is phosphorylated and activated by the constitutively active kinase of receptor II (Wrana et al., 1994). The activated type I receptor kinase then propagates the signal to a family of intracellular signaling mediators known as Smads (contraction of the C.elegans Sma and Drosophila Mad genes which were the first identified members of this class of signaling effectors).

Three classes of Smads with distinct functions have been defined: the receptor-regulated Smads, which include Smad1, 2, 3, 5, and 8; the common mediator Smad, Smad4; and the antagonistic Smads, which include Smad6 and 7 (Heldin et al., 1997; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Receptor-regulated Smads (R-Smads) act as direct substrates of specific type I receptors, and the proteins are phosphorylated on the last two serines at the carboxyl terminus within a highly conserved SSXS motif (Macías-Silva et al., 1996 ; Abdollah et al., 1997 ; Kretzschmar et al., 1997 ; Liu et al., 1997b ; Souchelnytskyi et al., 1997). Regulation of R-Smads by the receptor kinase provides an important level of specificity in this system. Thus, Smad2 and Smad3 are substrates of TGF β or activin receptors and mediate signaling by these ligands (Macías-Silva et al., 1996 ; Liu et al., 1997b ; Nakao et al., 1997), whereas Smad1, 5, and 8 are targets of BMP receptors and propagate BMP signals (Hoodless et al., 1996 ; Chen et al., 1997b ; Kretzschmar et al., 1997 ; Nishimura et al., 1998). Once phosphorylated, R-Smads associate with the common Smad, Smad4 (Lagna et al., 1996 ; Zhang et al., 1997), and mediate nuclear translocation of the heteromeric complex. In the nucleus, Smad complexes then activate specific genes through cooperative interactions with DNA and other DNA-binding proteins such as FAST1, FAST2, and Fos/Jun (Chen et al., 1996 , Chen et al., 1997a ; Liu et al., 1997a ; Labbé et al., 1998 ; Zhang et al., 1998 ; Zhou et al., 1998). In contrast to R-Smads and Smad4, the antagonistic Smads, Smad6 and 7, appear to function by blocking ligand-dependent signaling (reviewed in Heldin et al., 1997).

Phosphorylation of R-Smads by the type I receptor is essential for activating the TGF β signaling pathway (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). However, little is known of how Smad interaction with receptors is controlled. A novel Smad2/Smad3 interacting protein has been described (Tsukazaki T. et al., 1998) that contains a double zinc finger, or FYVE domain, and which has been called SARA (Smad anchor for receptor activation). The SARA motif recruits Smad2 into distinct subcellular domains and co-localizes and interacts with TGF β receptors. TGF β signaling induces dissociation of Smad2 from SARA with concomitant formation of Smad2/Smad4 complexes and nuclear translocation. Moreover, deletion of the FYVE domain in SARA causes mislocalization of Smad2 and inhibits TGF β -dependent transcriptional responses. Thus, SARA defines a component of TGF β signaling that functions to recruit Smad2 to the receptor by controlling the subcellular localization of Smad.

References: Abdollah et al. (1997) *J. Biol. Chem.* 272, 27678-27685; Attisano et al. (1998) *Curr. Opin. Cell Biol.* 10, 188-194; Chen et al. (1996) *Nature* 383, 691-696; Chen et al. (1997a) *Nature* 389, 85-89; Chen et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 12938-12943; Heldin et al. (1997) *Nature* 390, 465-471; Hoodless et al. (1996) *Cell* 85, 489-500; Kretschmar et al. (1998) *Curr. Opin. Genet. Dev.* 8, 103-111; Kretschmar et al. (1997) *Genes Dev.* 11, 984-995; Labbé et al. (1998) *Mol. Cell* 2, 109-120; Lagna et al. (1996) *Nature* 383, 832-836; Liu et al. (1997a) *Genes Dev.* 11, 3157-3167; Liu et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 10669-10764; Macías-Silva et al. (1996) *Cell* 87, 1215-1224; Nakao et al. (1997) *EMBO J.* 16, 5353-5362; Nishimura et al. (1998) *J. Biol. Chem.* 273, 1872-1879; Souchelnytskyi et al. (1997) *J. Biol. Chem.* 272, 28107-28115; Tsukazaki et al. (1998) *Cell* 95, 779-791; Wrana et al. (1994) *Nature* 370, 341-347; Zhang et al. (1997) *Curr. Biol.* 7, 270-276; Zhang et al. (1998) *Nature* 394, 909-913; Zhou et al. (1998) *Mol. Cell* 2, 121-127.

Calcium

The bivalent cation Ca^{2+} is, along with cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

Rab proteins

In eukaryotic cells the compartmentalization of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins and other molecules. Trafficking between organelles within the secretory pathway occurs as

vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle,

most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with α -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca^{2+} -binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca^{2+} influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits

that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991). *Mol. Cell. Biol.* 11, 872-885; Echard et al. (1998). *Science*. 279, 580-585; Geppert et al. (1998). *Annu. Rev. Neurosci.* 21, 75-95; Guo et al. (1999). *EMBO J.* 18, 1071-1080; Kato et al. (1996). *J. Biol. Chem.* 271, 31775-31778; Novick et al. (1997). *Curr. Opin. Cell Biol.* 9, 496-504; Peterson et al. (1999). *Curr. Biol.* 9, 159-162; Poirier et al. (1998). *Nat. Struct. Biol.* 5, 765-769; Vitale et al. (1998). *EMBO J.* 17, 1941-1951; Wang et al. (1997). *Nature*. 388, 593-598; Yang et al. (1999). *J. Biol. Chem.* 274, 5649-5653.

Kinases

Reversible posttranslational modifications of proteins are major means of regulating cellular activities. Among the various modifications that are carried out by the cells, the addition of phosphoryl groups to Ser/Thr or Tyr residues is the most important and widely used. The phosphorylation of proteins is accomplished by protein kinases, while the reverse reaction, the removal of phosphoryl groups, is carried out by phosphatases. Kinases / Phosphatases regulate key positions e.g. in the processes of cell proliferation, differentiation and communication/signaling. These processes must be tightly regulated in order to maintain a steady state level of cellular fate. Mis-regulation of kinase activities (or that of

phosphatases) is made responsible for a multitude of disease processes such as oncogenesis, inflammatory processes, arteriosclerosis, and psoriasis.

Protein kinases constitute the largest protein family that is currently known. Several hundred kinases have been identified already. Classically, kinases are subdivided into two classes based on the amino acid residues in their substrates that are phosphorylated by the particular enzymes. The kinases specifically add phosphoryl groups from adenosine triphosphate (ATP) or, less frequently, guanosine triphosphate (GTP), either to serine and/or threonine or to tyrosine residues of substrate proteins. An estimated 1,000 to 10,000 proteins present in a typical mammalian cell are believed to be regulated also by the action of protein kinases.

Protein kinases are frequently integral parts of signaling cascades that transmit extracellular stimuli (e.g. hormones, neurotransmitters, growth- or differentiation factors) into the cell and result in various responses by the cells. The kinases play key roles in these cascades as they constitute a sort of 'molecular switches' turning on or off the activities of other enzymes and proteins, e.g. metabolic, regulatory, channels and pumps, receptors, cytoskeletal, transcription factors.

The regulation of kinase activities is accomplished by various means:

The best characterized example for the regulation via regulatory subunits is the cAMP-dependent protein kinase (PKA) which is also a prototype for second messenger activated protein kinases. This enzyme consists of a heterotetramer of two catalytic (C) and two regulatory (R) subunits. Upon binding of two molecules of second messenger (cAMP) in each R subunit, the catalytic subunits are released and active. Both of the catalytic and the regulatory subunits several isoforms exist. The combination of catalytic and regulatory subunits determines the localization of the holoenzyme and also the substrate spectrum that is available for phosphorylation. The consensus pattern necessary to be present in the substrate for PKA action is RRXS/T where X can be any amino acid.

The casein kinase II comprises another examples for holoenzymes that consist of catalytic and regulatory subunits. Other kinases that are activated by second messengers are cGMP-dependent protein kinase and Protein kinase C (PKC) which is activated by diacylglycerol, which in turn is produced by phospholipases by cleavage of phosphatidylcholine.

Receptor kinases usually consists of an extracellular domain which can bind effector molecules (e.g. growth factors and hormones) and transfer the stimulus to the intracellular domain of these proteins which usually is a protein tyrosine kinase. Other tyrosine kinases lack an extracellular domain but are associated with receptors which transfer the signal after effector binding by activating the associated protein kinase enzyme (e.g. Src kinase family; Src, Blk, Fgr, Fyn, Lck Lyn, Yes and Janus kinase family; Jak1-3, Tyk2).

Dysfunction of kinases, e.g. caused by non-functioning regulation, can be the cause of inflammatory diseases and uncontrolled proliferation. v-Src which is a truncated version of the C-Src protooncogene tyrosine kinase is a classical example for this process as v-Src does not contain the regulatory domain of the cellular gene and is thus constitutively active.

Several categories of proteins are coded for by clones of the invention within the overall group of "Signal transduction" and include, among others, the following:

Neurocalcin (Recoverin): Neurocalcin is a Ca^{2+} -binding protein with three putative Ca^{2+} -binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca^{2+} dependent activation of guanylate cyclase.. These proteins can find application in modulating/blocking the guanylate cyclase-pathway. Diseases associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) autosomal dominant cone dystrophy (OMIN *600364); 2) cone dystrophy 3 (OMIN *600364); 3) cancer associated retinopathy (OMIN *179618). Clones in this category include: fbr2_23b21.

Proteins with a WW Domain: Proteins that contain a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore these proteins should be involved in intracellular signal transduction. Diseases associated (as

potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200). Clones in this category include: fbr2_23n16.

Protein substrates for cAMP-dependent protein kinase: Acting as a choride channel or chloride channel inhibitor these proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Cystic Fibrosis (OMIN #219700). Clones in this category include fbr2_82i17.

Sphingosine kinase: Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellularly, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependent on SPP. Extracellularly, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1. These proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Gaucher Disease, Type I (OMIN *230800). Clones in this category include fbr2_82m6.

Vanilloid Receptors: VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. Related can find application as a target for the development of new nociception-modulating drugs. Clones in this category include tes3_20k2.

RCC1 (Regulator of chromosome condensation): RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator. These proteins can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat. OMIN also reports that RCC1 has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with retinitis pigmentosa (OMIN *312610). Clones in this category include tes3_21d4.

Ras inhibitor proteins: Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show

intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. Ras inhibitor proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with many disease processes as reported by OMIN including: 1) Tumors of the lung, breast, brain, pituitary, pancreas, bone, skin, bladder, kidney, ovary, prostate and lymphocyte, Melanoma (OMIN *600160); 2) X-linked non-specific mental retardation (OMIN *300104); 3) adenomatous polyposis of the colon (OMIN *175100); 4) Beckwith-Wiedemann Syndrome (#130650); and 5) Major affective disorder 1 (OMIN *125480). Clones in this category include ute1_22g21.

Mammalian proteins cornicon involving the EGF-receptor: Cornicon proteins are part of a signal transduction pathway involving the EGF-receptor. The EGF-receptor has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Leprechaunism (OMIN #246200); 3) Hemophilia B (OMIN *306900); 4) Ectodermal dysplasia 1; 5) Kartagener syndrome (OMIN *244400) and 6) Glioma of the brain (OMIN *137800). Clones in this category include ute1_22e12.

Transmembrane proteins

Membrane region prediction was effected using the ALOM2 software (Klein et al., 1985; version 2 by K. Nakai). Similar to many other methods, the Kyte & Doolittle (1982) amino acid hydrophobicity scale is used in ALOM2 as the primary variable for classifying sequences in terms of their localization. High prediction accuracy is achieved through the system of intelligent decision rules and the utilization of a carefully selected training data set. The method also generates reliability estimates which makes it possible to distinguish between membrane-spanning proteins (I, intrinsic) and globular proteins with regions of high hydrophobicity buried in the core.

For a protein of length L , the block of length l with maximum hydrophobicity is found:

$$\max H = \max_{k=1, \dots, L-l+1} (1/l) \sum_{i=k}^{k+l-1} H_i$$

where H_i represents the hydrophobicity of an individual residue.

Let $P(I/\max H)$ and $P(E/\max H)$ be the conditional probabilities that a protein is integral or peripheral, respectively, given its value of maximal hydrophobicity $\max H$, and let $P(I)$ and $P(E)$ be the prior probabilities of intrinsic and extrinsic membrane proteins estimated from the training set. Then a sequence is assigned to E if

$$P(E/\max H) > P(I/\max H)$$

or, after applying the Bayes rule,

$$P(E)P(\max H/E) > P(I)P(\max H/I),$$

where the conditional probabilities $P(\max H/E)$ and $P(\max H/I)$ can be determined based on the estimates of probability distributions of $\max H$ in both groups.

Discriminant analysis allows to simplify this task by calculating the odds $P(E/\max H):P(I/\max H)$ as e^b , where b is the left-hand side of a linear or quadratic inequality. For example, for the window of length 17, the protein is allocated to the peripheral category E based on the empirically derived quadratic inequality:

$$1.05(\max H)^2 + 12.30\max H + 17.49 > 0,$$

whereas the optimal inequality for assigning membrane proteins (category I) is linear:

$$-9.02\max H + 14.27 > 0$$

The odds parameter can be made more or less stringent. For example, one can require odds at least 1:10 for a protein to be classified as integral. This leads to higher selectivity but less sensitivity.

The boundaries of membrane-spanning regions in putative membrane proteins are detected by means of an iterative procedure whereby the most hydrophobic region corresponding to the value $\max H$ is considered to be membrane and removed from the sequence. The classification procedure is then repeated again for the remaining sequence, and, if such a protein is again classified as integral, the next most hydrophobic region is considered.

Reference: Klein, P., Kanehisa, M., DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochem Biophys Acta* **815**: 468-476

Transcription factors

Purified eukaryotic RNA polymerase II is unable to initiate promoter-specific transcription. A family of factors that collectively confer RNAPII promoter specificity is known as the general transcription factors (GTFs). They include the TATA-binding Protein (TBP) TFIIB, TFIIE, TFIIIF and TFI IH. These factors are conserved among all eukaryotes.

RNAPII complexes containing the entire set of GTFs or a subset of GTFs together with other proteins have been isolated from mammalian and yeast cells. Although purified RNAPII and GTFs are sufficient for promoter-specific initiation, this system fails to respond to activators. This is mediated by a further complex termed mediator complex which associates with the carboxy-terminal heptapeptide domain (CTD) of the largest subunit of RNAPII.

Purification of human RNAPII complexes resulted in two distinct forms of human RNAPII after analysis of functional properties. One complex contained chromatin remodeling activities but was devoid of GTFs. The other complex did not contain factors that modify chromatin but contained a subset of SRB/mediator subunits and GTFs and other polypeptides that mediate transcriptional activation, a scenario similar to that reported for yeast.

A complex designated NAT (~20 SU) for negative regulator of transcription contains RNAPII, Cdk8, homologs of the yeast mediator complex as well as Rgr1 and Srb10/11 known as negative regulators of transcription.

A complex with striking similar structural and functional properties to NAT has been identified designated SMCC (~15 SU) (SRB/mediator coactivator complex), that can also mediate transcriptional activation.

The SMCC complex includes all reported NAT subunits including subunits of the TRAP complex. TRAP is a coactivator complex isolated on the basis of its interaction with the thyroid hormone receptor. Another coactivator complex DRIP, isolated on the basis of its

ability to interact with the vitamin D3 receptor, contains novel subunits as well as subunits of NAT/SMCC and TRAP complexes.

The effects of each of these coactivator complexes is dependent on the TFIID complex. It is not known if the TAF subunits of TFIID are required. It is likely that new coactivator complexes will be uncovered containing both novel and previously defined components.

Beside the huge amount of transcription factors which can be part of the RNAIIP holoenzyme or the coactivator complexes there is an even larger quantity of specific transcription factors binding to promoter elements within the DNA sequences of a given gene leading to activation or repression of transcription. A broad range of cellular responses like differentiation, proliferation, cell death and others are elicited through activating or repressing the transcription of target genes.

There are at least five superclasses of transcription factors:

1. Superclass contains members with characteristic basic domains:

Members are:

Leucine zipper factors, where the basic domain is followed by a leucine zipper of repeated leucine residues at every seventh position. The zipper mediates protein dimerization as a prerequisite for DNA-binding.

Helix-loop-helix factors (bHLH) contain a DNA-binding basic region followed by a motif of two potential amphipathic alpha-helices connected by a loop of variable length also mediating dimerization.

Factors with a combination of Helix-loop-helix and leucine zipper.

Further members of this superclass are NF- κ B, RF-X, and bHLH like proteins.

2. Superclass comprises factors containing zinc-coordinating DNA-binding domains.

Members are:

Proteins with Cys4 zinc finger of nuclear receptor type, where two such motifs differing in size, composition and function are present in each receptor molecule. Each finger comprises 4 cysteine residues coordinating one zinc ion. The second half including the second cysteine pair has alpha-helix conformation and the helix of the first finger binds to the DNA through the major groove. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding. This class includes the steroid hormone receptors and the thyroid hormone receptor-like factors. Other diverse cys4 zinc fingers have a motif of GATA-type.

Proteins with Cys2His2 zinc finger domain(s). Each finger comprises 2 cysteine and 2 histidine residues coordinating one zinc ion, and in some cases one histidine is replaced by another cysteine. The zinc ion is essential for DNA-binding.

Proteins with Cys6 cysteine-zinc cluster(s). Six cysteine residues coordinate two zinc ions, i. e. two of the thiol groups are coordinating two zinc ions each. Present in many fungal regulators.

Zinc fingers of alternating composition.

3. Superclass contains factors of helix-turn-helix type.

Members are:

Proteins with homeo domains. Homeo domains are three consecutive alpha-helix structures. Helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.

Proteins with Paired box domain(s). This is a DNA-binding domain of approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half is highly charged in general. It probably comprises 3 alpha-helices.

Proteins with Fork head / winged helix domain(s). This domain was identified by homology between HNF-3A and fkh. The domain comprises approx. 110 AA. Analysis of the crystal structure has revealed a compact structure of three alpha-helices, the third alpha-helix

being exposed towards the major groove of the DNA. The domain also exerts minor groove contacts. Upon binding to DNA, it induces a bend of 13 degree.

Heat shock factors

Proteins with Tryptophan clusters. The tryptophan clusters comprise several tryptophan residues with a spacing of 12-21 amino acid residues; the subclass of myb-type DNA-binding domains typically exhibit a spacing of 19-21 amino acid residues.

Proteins with TEA domain(s). The TEA domain has been identified as a region which is conserved among the transcription factors TEF-1, TECl and abaA. This domain in TEF-1 has been shown to interact with DNA, although two additional regions may also contribute to DNA-binding. It is predicted to fold into three alpha-helices, with a randomly coiled region of 16-18 amino acid residues between helices 1 and 2, and a short stretch between helices 2 and 3 of 3-8 residues.

4. Superclass contains beta-Scaffold Factors with Minor Groove Contacts

Members are:

Proteins with RHR (Rel homology) region.

The structure of the Rel-type DBD exhibits a bipartite subdomain structure, each subdomain comprising a beta-barrel with five loops that form an extensive contact surface to the major groove of the DNA. Particularly, the first loop of the N-terminal subdomain (the highly conserved recognition loop) performs contacts with the recognition element on the DNA, but other loops are involved. The fact that the main DNA-contacts are made through loops has been suggested to provide a high degree of flexibility in binding to a range of different target sequences. Augmenting interactions are achieved by two alpha-helices within the N-terminal Part that form strong minor groove contacts to the A/T-rich center of the B-element. In p65, the sequence between both alpha-helices is much shorter and even helix 2 is truncated. The second, C-terminal domain is necessary mainly for protein dimerization.

p53 proteins

MADS (MCM1-agamous-deficiens-SRF) box proteins. Proteins of this class comprise a region of homology. The DNA-binding domain also comprises the dimerization capability. In the DNA-bound dimer (shown for SRF), two antiparallel amphipathic alpha-helices (alpha-I), form a coiled coil and are oriented approximately parallel on the minor groove. These helices make minor and major groove contacts, the N-terminal extensions form minor groove contacts. The bound DNA is bent and wrapped around the protein. It exhibits a compressed minor groove in the center and widened minor groove in the flanks.

Beta-Barrel alpha-helix transcription factors.

TATA-binding proteins

HMG proteins

Proteins of this class comprise a region of homology with the chromosomal non-histone HMG proteins such as HMG1. This region comprises the DNA-binding domain which in some instances such as HMG1 mediates sequence-unspecific, in other cases such as LEF-1 sequence-specific binding to DNA. This domain exhibits a typical L-shaped conformation made up of 3 alpha-helices and an extended N-terminal extension of the first helix. The latter together with helix 1, which contains a kink, form the long arm of the L, whereas helices 1 and 2 form the short arm. Binding to the minor groove induces a sharp bending of the DNA by more than 90 degree, away from the bound protein. The overall topology of the DNA-protein complexes resembles somewhat that of the TBP-TATA box complex.

Heteromeric CCAAT factors

Proteins with Grainyhead domain(s)

Cold-shock domain factors. Cold-shock domain proteins are characterized by a highly conserved region first found in prokaryotic cold-shock proteins. This domain is a single-stranded nucleic acid-binding structure interacting with DNA or RNA. It consists of an antiparallel five-stranded beta-barrel, the strands of which are connected by turns and loops. Within this structure, a three-stranded beta-strand contains a conserved RNA-binding motif, RNPI. Not all CSD proteins are transcription factors. Those which specifically bind to a

certain sequence are termed Y-box proteins. Proteins of this class were previously called protamine-like domain proteins because of having a highly positively charged domain with interspersed proline residues.

Proteins with Runt homology domain

The members of this transcription factor class have been identified on the basis of their homology to a defined region within the *Drosophila* protein Runt. The runt domain is part of the DNA-binding domain of these factors. It consists mainly of beta-strands, does not contain alpha-helical regions and seems to be most similar to the palm domain found in DNA polymerase beta (rat).

5. Superclass contains other transcription factors like Copper fist proteins, HMGI(Y), STAT, Pocket domain proteins and Ap2/EREBP-related factors.

The classification of transcription factors originates from TRANSFAC database:

<http://transfac.gbf.de/TRANSFAC/>

Reference: Heinemeyer

Several categories of proteins are coded for by clones of the invention within the overall group of "Transcription Factors" and include, among others, the following:

Dcoh: Dcoh is a bifunctional protein, complexed with bioppterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bioppterin cofactor of phenylalanine hydroxylase. The Dcoh protein has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hyperphenylalanemia (OMIN 126090, #264070). Clones in this category include fkd2_46k12.

Signal transducing proteins: Beta-transducin subunits of G-proteins contain WD-40 repeats. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription. These proteins have been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) essential hypertension (OMIN *139130). Clones in this category include ute1_1i2.

* * *

The invention, therefore, specifically contemplates the following assemblages of materials, which track the above-identified fourteen functional groupings, that are useful in practicing the profiling aspects of the invention. One type of assemblage is nucleic acid-based and can include the following groupings of sequences and their derivatives: all sequences; human fetal brain sequences; brain derived sequences; human fetal kidney library sequences; kidney derived sequences; human mammary carcinoma library sequences; mammary carcinoma derived sequences; human testis library sequences; testes derived sequences; cell cycle genes; cell structure and motility genes; differentiation and development genes; intracellular transport and trafficking genes; metabolism genes; nucleic acid management genes; signal transduction genes; transmembrane protein genes; and transcription factor genes. Other assemblages contain proteins or their corresponding antibodies or antibody fragments, divided along the same groupings.

Database Applications

Because they are human genes and gene products, the inventive molecules are useful as members of a database. Such a database may be used, for example, in drug discovery and rationale drug design or in testing the novelty and non-obviousness of newly sequenced materials. In addition, they are particularly suited in designing variants for the profiling (and other) applications described herein. Hence, the following discussion of electronic embodiments applies equally to such variants, which, naturally, will be generated and stored using a computer using known methodologies.

Accordingly, one aspect of the invention contemplates a database of at least one of the inventive sequences stored on computer readable media. Again, the individual sequences may be grouped with regard to the individual functional and structural groups mentioned above. While the individual sequences of a database may exist in printed form, they are preferably in electronic form, as in an ascii or a text file. They may also exist as word processing files or they may be stored in database applications like DB2, Sybase, Oracle, GCG and GenBank. One skilled in the art will understand the range of applications suitable for using and storing the electronic embodiments of the invention.

"Computer readable media" refers to any medium which can be read and accessed by a computer. These include: magnetic storage media, like floppy discs, hard drives and magnetic tape; optical storage media, like CD-ROM; electrical storage media, like RAM

and ROM; and hybrids of these categories, like magnetic/optical storage media. One skilled in the art will readily understand the scope of computer readable media and how to implement them.

Biological Activities and Assays for Implementing Therapeutic and Diagnostic Applications

This section provides assays for biological activity that are useful in characterizing and quantifying the biological activity of the inventive molecules and their derivatives, which is relevant to the pharmacological effects of the inventive molecules. As used in this section, it will be understood that "protein" may also refer to the inventive antibodies (including fragments).

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M + (preB M +), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin gamma, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6-Nordan, R. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to modify immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the

tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this manner prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-vital immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient.

The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and beta 2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Bowman et al., *J. Virology* 61:1992-1998; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnolli et al., *Cellular Immunology* 133:327-341, 1991; Brown et al., *J. Immunol.* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; and Assays for B cell function: *In vitro* antibody production, Mond, J. J. and Brunswick, M. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnolli et al., *J. Immunol.* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., *J. Immunol.* 134:536-544, 1995; Inaba et al., *Journal of*

Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad. Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelosuppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the

treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and

cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle

stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin alpha family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- beta group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of

cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such

receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of

cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in

a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Particular Applications for Certain Clones

The following sets out a non-exclusive list of applications for certain embodiments of the invention. In the interest of economy, applications relevant to multiple embodiments are not duplicated in this list. Other embodiments described in below have similar characteristics, as described therein. The artisan is directed, therefore, to this section for similar descriptions of the functions of other embodiment.

Testes

htes3_15c24: The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

htes3_15i5: The new protein can find application in modulating the structure of the human spermatozoa radia spoke head and modulation of sperm motility in men.

htes3_15k11: The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

htes3_17n12: The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

htes3_20k2: The new protein can find application as a target for the development of new nociception-modulating drugs.

htes3_20m18: The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

htes3_20d4: The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

htes3_21j15: NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor. The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

htes3_26g22: The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP. The new protein can find application in modulation of gene transcription.

htes3_21l16: The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

htes3_27d1: The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

htes3_2m18: The novel protein can find application as multifunctional nuclease / exoribonuclease.

htes3_35b4: The new protein can find application in modulation of the mitotic spindle.

htes3_35b5: The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

htes3_35e21: Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

htes3_35k16: Therefore it is a new fatty acid-CoA synthetases/ligase with unknown substrate. The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

htes3_35n12: The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

htes3_35n9: The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

htes3_35p22: The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control. The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

htes3_4h6: The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

htes3_72k15: FGD1-related F-actin-binding protein (Frabin/FGD1) is a novel F-actin-binding protein. The gene locus *fgd1* seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

htes3_72p16: As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP). The new protein can find application in modulation the sorting of proteins into different compartments.

htes3_7b22: The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

htes3_7j3: The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too. The new protein can find application in modulating/blocking the cell cycle.

htes3_7p9: The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle. The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex. The new protein can find application in modulation of viral infections and tumour events.

htes3_8m10: The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA. The new protein can find application in modulation of mRNA translation and processing/stability.

Kidney

hfk2_24b15: The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

hfkd2_24n20: The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

hfkd2_3o17: The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

hfkd2_46j20: The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

hfkd2_46k19: The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

hfkd2_46m4: SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

hfkd2_46k14: rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport. The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

Uterus Associated:

hutel_18i19: The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

hutel_18l1: The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome sub-unit.

hutel_19g22: The new protein can find application in modulation of tissue-calcification, especially the uterus.

hutel_19h17: The new protein can find application in modulating the response of cells to oxysterols.

hutel_20b19: The novel protein seems to be a novel enzyme with sarcosine oxidase activity. The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

hutel_20g21: The novel protein seems to be a new ras inhibitor protein. The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

hutel_20h13: The novel protein is a new human alpha-adaptin. The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

hutel_20m11: The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

hutel_20m24: This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2. The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

hutel_22e12: The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

hutel_23e13: The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

hutel_24j6: The new protein can find application in modulation of cell-cell-adhesion.

hutel_24h3: The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

Fetal Brain:

hfbr2_16c16: The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

hfbr2_23b21: The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

hfbr2_23b10: The new protein can find application in modulation of splicing.

hfbr2_2b5: The novel protein contains the typical (xxG)_n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain. The new protein can find application in modulation of connective tissue, bone and cartilage development and maintainance.

hfbr2_2c17: The new protein can find application in modulating/blocking G-protein-dependent pathways.

hfbr2_2d15: The new protein can find application in modulating early spermatogenesis.

hfbr2_2i17: The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

hfbr2_2k14: Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

hfbr_3c18: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

hfbr_3g8: The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

hfbr2_62b11: The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

hfbr2_62o17: The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins.

hfbr_6b24: The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

hfbr_72b18: The new protein can find application in modulating DNA repair and mutagenesis.

hfbr_78c4: The new protein can find application in modulating/blocking the response of cells to interferons.

hfbr_78k24: These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. The new protein can find application in modulation of protein stability/degradation in cells.

hfbr_82e4: The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

VARIANTS OF THE INVENTIVE DNA MOLECULES

Variants in General

"Variants," according to the invention, include DNA and/or protein molecules that resemble, structurally and/or functionally, those set forth in herein. Variants may be isolated from natural sources ("homologs"), may be entirely synthetic or may be based in part on both natural and synthetic approaches.

The section set forth below presents various structural and functional characteristics of molecules within the invention. Preferred molecules are characterized by a combination of one or more of these characteristics. For instance, some preferred molecules are described with reference to at least two structural characteristics, while others may be described with reference to at least one structural and at least one functional characteristic.

It will be recognized by the skilled artisan that structure ultimately defines function, *i.e.* the functions of the molecules described herein derives from the structures of those

molecules. Accordingly, the structural variants described below that bear the closest structural relationship (as variously defined below) to the inventive molecules are the variants that most likely will preserve biological function. This relationship between structure and function will guide the skilled artisan in identifying the preferred embodiments of the invention.

Splicing Variants

It is well-known that eukaryotic structural genes are comprised of both protein coding and non-coding portions. When the messenger RNA is transcribed from the DNA template, it contains introns, which are non-coding, and exons, which are coding. In order to form a translation competent mRNA, the introns must be "spliced" out of this initial pre mRNA.

Specific sequences within the pre mRNA represent "splice junctions" that direct the cellular splicing machinery to the appropriate position. The splice junctions are loosely conserved sequence regions of the pre mRNA, which almost invariably begin with GT and end with AG (DNA perspective). The 5' end of the splice junction typically contains about nine somewhat conserved residues, for example, C/AAGTA/GAGT. The 3' end usually contains a pyrimidine rich stretch of at least about 11 nucleotides, followed by NC/TAGG. Splicing occurs before the GT and after the AG. Mount, *Nucleic Acids Res.* 10:459-72 (1982).

Interestingly, exons often correspond to discrete functional domains of the protein product. The intron/exon arrangement thus creates a linear array of nucleotides which can be correlated to discrete, and often interchangeable, functional protein fragments. Go, *Nature* 291:90-92 (1981); Branden *et al.*, *EMBO J.* 3:1307-10 (1984). This linear arrangement creates the possibility of generating multiple different full length proteins by rearranging the order of the different functional portions in the array. For example, if a set of exons are arranged 1-2-3-4, where (-) represents the introns separating the exons, a splicing event need not simply produce 1234, but may produce 123, 134, 124 and so on. Production of different mRNA products in this way is commonly called "alternative splicing." Andreadis *et al.*, *Ann. Rev. Cell Biol.* 3:207-42 (1987).

Some of the present DNA molecules can be represented in modular fashion in terms of their coding regions. Essentially, these modules are exons (though each "exon" may in fact be made up of several exons), which may be combined in different ways to form a variety of

different DNA molecules, each encoding a different functional protein. Splicing variants are indicated below.

Degenerate Variants

One aspect of the present invention provides "degenerate variants" of the nucleic acid fragments of the present invention. A "degenerate variant" is a nucleotide fragment which differs from those of inventive molecules by nucleotide sequence, but due to the degeneracy of the genetic code, encodes an identical polypeptide sequence.

Given the known relationship between DNA sequences and the proteins they encode, degenerate variants typically are described by reference to this relationship. It is well known that the degeneracy of the genetic code results in many possible DNA sequences which encode a particular protein. Indeed, of the three bases which comprise an amino acid-encoding triplet, the third position, and often the second, almost always may vary. This fact alone allows for a class of variant DNA molecules which encode protein sequences identical to those disclosed herein, yet have about 30% sequence variation. In other words, the variant DNA molecules are about 70% identical to the inventive DNAs, having no additional or deleted sequences. Thus, one aspect of the invention provides degenerate variant DNA molecules encoding the inventive protein sequences.

In one embodiment, these variants have at least about 70% sequence identity with the DNA molecules described herein. In a preferred embodiment, these variants have at least about 80% sequence identity to the inventive molecules. In a more preferred embodiment these variants have at least about 90% sequence identity with the inventive molecules.

Conservative Amino Acid Variants

Variants according to the invention also may be made that conserve the overall molecular structure of the encoded proteins. Given the properties of the individual amino acids comprising the disclosed protein products, some rational substitutions will be recognized by the skilled worker. Amino acid substitutions, *i.e.* "conservative substitutions," may be made, for instance, on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example: (a) nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; (b) polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine;

(c) positively charged (basic) amino acids include arginine, lysine, and histidine; and (d) negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Substitutions typically may be made within groups (a)-(d). In addition, glycine and proline may be substituted for one another based on their ability to disrupt α -helices. Similarly, certain amino acids, such as alanine, cysteine, leucine, methionine, glutamic acid, glutamine, histidine and lysine are more commonly found in α -helices, while valine, isoleucine, phenylalanine, tyrosine, tryptophan and threonine are more commonly found in β -pleated sheets. Glycine, serine, aspartic acid, asparagine, and proline are commonly found in turns. Some preferred substitutions may be made among the following groups: (i) S and T; (ii) P and G; and (iii) A, V, L and I. Given the known genetic code, and recombinant and synthetic DNA techniques, the skilled scientist readily can construct DNAs encoding the conservative amino acid variants.

As used herein, "sequence identity" between two polypeptide sequences indicates the percentage of amino acids that are identical between the sequences. "Sequence similarity" indicates the percentage of amino acids that either are identical or that represent conservative amino acid substitutions.

Functionally Equivalent Variants

Yet another class of DNA variants within the scope of the invention may be described with reference to the product they encode. As shown below, some of the inventive DNA molecules encode a protein having a degree of homology with known proteins, or protein domains. It is expected, therefore, that they will have some or all of the requisite functional features of such molecules. These "functionally equivalent variants" products are characterized by the fact that they are functionally equivalent, with respect to biological activity, to certain known molecules.

The instant invention provides information on common structural motifs, including consensus sequences that will guide the artisan in constructing functionally equivalent variants. It will be understood that the motifs, identified for each inventive protein, may be modified within the identified consensus sequences. Thus, the invention contemplates the proteins disclosed herein that contain variability in the consensus sequences identified, and the invention further contemplates the full range of nucleic acids encoding them, and the complements of those nucleic acids.

Hybridizing Variants

DNA variants within the invention also may be described by reference to their physical properties in hybridization. One skilled in the field will recognize that DNA can be used to identify its complement and, since DNA is double stranded, its equivalent or homolog, using nucleic acid hybridization techniques. It will also be recognized that hybridization can occur with less than 100% complementarity. However, given appropriate choice of conditions, hybridization techniques can be used to differentiate among DNA sequences based on their structural relatedness to a particular probe. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, N.Y.; and Ausubel *et al.*, 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Green Publishing Associates and Wiley Interscience, N.Y.

Structural relatedness between two polynucleotide sequences can be expressed as a function of "stringency" of the conditions under which the two sequences will hybridize with one another. As used herein, the term "stringency" refers to the extent that the conditions disfavor hybridization. Stringent conditions strongly disfavor hybridization, and only the most structurally related molecules will hybridize to one another under such conditions. Conversely, non-stringent conditions favor hybridization of molecules displaying a lesser degree of structural relatedness. Hybridization stringency, therefore, directly correlates with the structural relationships of two nucleic acid sequences. The following relationships are useful in correlating hybridization and relatedness (where T_m is the melting temperature of a nucleic acid duplex):

- a. $T_m = 69.3 + 0.41(G+C)\%$
- b. The T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched base pairs.
- c. $(T_m)_{\mu 2} - (T_m)_{\mu 1} = 18.5 \log_{10} \mu 2 / \mu 1$
where $\mu 1$ and $\mu 2$ are the ionic strengths of two solutions.

Hybridization stringency is a function of many factors, including overall DNA concentration, ionic strength, temperature, probe size and the presence of agents which disrupt hydrogen bonding. Factors promoting hybridization include high DNA

concentrations, high ionic strengths, low temperatures, longer probe size and the absence of agents that disrupt hydrogen bonding.

Hybridization usually is done in two stages. First, in the "binding" stage, the probe is bound to the target under conditions favoring hybridization. Stringency is usually controlled at this stage by altering the temperature. For high stringency, the temperature is usually between 65°C and 70°C, unless short (<20 nt) oligonucleotide probes are used. A representative hybridization solution comprises 6X SSC, 0.5% SDS, 5X Denhardt's solution and 100µg of non-specific carrier DNA. See Ausubel *et al.*, *supra*, section 2.9, supplement 27 (1994). Of course many different, yet functionally equivalent, buffer conditions are known. Where the degree of relatedness is lower, a lower temperature may be chosen. Low stringency binding temperatures are between about 25°C and 40°C. Medium stringency is between at least about 40°C to less than about 65°C. High stringency is at least about 65°C.

Second, the excess probe is removed by washing. It is at this stage that more stringent conditions usually are applied. Hence, it is this "washing" stage that is most important in determining relatedness via hybridization. Washing solutions typically contain lower salt concentrations. One exemplary medium stringency solution contains 2X SSC and 0.1% SDS. A high stringency wash solution contains the equivalent (in ionic strength) of less than about 0.2X SSC, with a preferred stringent solution containing about 0.1X SSC. The temperatures associated with various stringencies are the same as discussed above for "binding." The washing solution also typically is replaced a number of times during washing. For example, typical high stringency washing conditions comprise washing twice for 30 minutes at 55° C. and three times for 15 minutes at 60° C.

The present invention includes nucleic acid molecules that hybridize to the inventive molecules under high stringency binding and washing conditions. More preferred molecules (from an mRNA perspective) are those that are at least 50 % of the length of any one of those depicted in below. Particularly preferred molecules are at least 75 % of the length of those molecules.

Substitutions, Insertions, Additions and Deletions

In a general sense, the preferred DNA variants of the invention are those that retain the closest relationship, as described by "sequence identity" to the inventive DNA molecules. According to another aspect of the invention, therefore, substitutions, insertions, additions and deletions of defined properties are contemplated. It will be recognized that sequence

identity between two polynucleotide sequences, as defined herein, generally is determined with reference to the protein coding region of the sequences. Thus, this definition does not at all limit the amount of DNA, such as vector DNA, that may be attached to the molecules described herein. Preferred DNA sequence variants include molecules encoding proteins sharing some or all of any relevant biological activity of the native molecule.

In creating these variants, the skilled worker will be guided by reference to the protein structure. First, insertions and deletions in any recognized functional domain, above, generally should be avoided, except as noted below in the section entitled "Proteins," where this domain is discussed in detail. Alterations in such domains usually will be limited to conservative amino acid substitutions. In addition, where insertions and deletions are desired, this may be accomplished at the N- and/or C-terminus of the protein molecule (or the corresponding coding regions of the DNA). If insertions or deletions are made within the protein, deletions of major structural features usually should be avoided. Thus, a preferred place to make insertion or deletion variants is in non-structural regions, such as linker regions between two alpha helices.

"Substitutions" generally refer to alterations in the DNA sequence which do not change its overall length, but only alter one or more nucleotide positions, substituting one for another in the common sense of the word. One class of preferred substitutions, "degenerate substitutions," are those that do not alter the encoded amino acid sequence. Some substitutions retains 50%, 55%, 60% or 65% identity. Preferred substitutions retain at least about 70% identity, more preferably at least 70% or 75% identity, with the inventive DNAs. Some more preferred molecules have at least about 80% identity, more preferably at least 80% or 85% identity. Particularly preferred DNAs share at least about 90% identity, more preferably at least 90% or 95% identity.

"Insertions," unlike substitutions, alter the overall length of the DNA molecule, and thus sometimes the encoded protein. Insertions add extra nucleotides to the interior (not the 5' or 3' ends) of the subject DNAs. Preferred insertions are made with reference to the protein sequence encoded by the DNA. Thus, it is most preferred to provide an insertion in the DNA at a location that corresponds to an area of the encoded protein which lacks structure. For instance, it typically would not be beneficial, if the preservation of biological activity is desired, to provide an insertion within an alpha-helical region or a beta-pleated sheet. Accordingly, non-structural areas, such as those containing helix-breaking glycines

and proline residues, are most preferred sites of insertion. Other preferred sites of insertion are the splice sites, which are indicated above in the description of the inventive DNA molecules.

While the optimal size of insertions will vary depending upon the site of insertion and its effect on the overall conformation of the encoded protein, some general guides are useful. Generally, the total insertions (irrespective of their number) should not add more than about 30% (or preferably not more than 30%) to the overall size of the encoded protein. More preferably, the insertion adds less than about 10-20% (yet more preferably 10-20%) in size, with less than about 10% being most preferred. The number of insertions is limited only by the number of suitable insertions sites, and secondarily by the foregoing size preferences.

"Additions," like insertions, also add to the overall size of the DNA molecule, and usually the encoded protein. However, instead of being made within the molecule, they are made on the 5' or 3' end, usually corresponding to the N- or C- terminus of the encoded protein. Unlike deletions, additions are not very size-dependent. Indeed, additions may be of virtually any size. Preferred additions, however, do not exceed about 100% of the size of the native molecule. More preferably, they add less than about 60 to 30% to the overall size, with less than about 30% being most preferred.

"Deletions" diminish the overall size of the DNA and, therefore, also reduce the size of the protein encoded by that DNA. Deletions may be made from either end of the molecule or internal to it. Typical preferred deletions remove discrete structural features of the encoded protein. For example, some deletions will comprise the deletion of one or more exons which may define a structural feature. Preferred deletions remove less than about 30% of the size of the subject molecule. More preferred deletions remove less than about 20% and most preferred deletions remove less than about 10%.

Computer-Defined Variants and Definition of "Sequence Identity"

In general, both the DNA and protein molecules of the invention can be defined with reference to "sequence identity." As used herein, "sequence identity" refers to a comparison made between two molecules using, for example, the standard Smith-Waterman algorithm that is well known in the art.

Some molecules have at least about 50%, 55% or 60% identity. Preferred molecules are those having at least about 65% sequence identity, more preferably at least 65% or 70% sequence identity. Other preferred molecules have at least about 80%, more preferably at

least 80% or 85%, sequence identity. Particularly preferred molecules have at least about 90% sequence identity, more preferably at least 90% sequence identity. Most preferred molecules have at least about 95%, more preferably at least 95%, sequence identity. As used herein, two nucleic acid molecules or proteins are said to "share significant sequence identity" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) identity.

"Sequence identity" is defined herein with reference the Blast 2 algorithm, which is available at the NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>), using default parameters.

References pertaining to this algorithm include: those found at

http://www.ncbi.nlm.nih.gov/BLAST/blast_references.html; Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266-272; Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131-141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402; and Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656.

METHODS OF MAKING VARIANTS

It will be recognized that variants of the inventive molecules can be constructed in several different ways. For example, they may be constructed as completely synthetic DNAs. Methods of efficiently synthesizing oligonucleotides in the range of 20 to about 150 nucleotides are widely available. See Ausubel *et al.*, *supra*, section 2.11, Supplement 21 (1993). Overlapping oligonucleotides may be synthesized and assembled in a fashion first reported by Khorana *et al.*, J. Mol. Biol. 72:209-217 (1971); see also Ausubel *et al.*, Section 8.2. The synthetic DNAs are designed with convenient restriction sites engineered at the 5' and 3' ends of the gene to facilitate cloning into an appropriate vector.

An alternative method of generating variants is to start with one of the inventive DNAs and then to conduct site-directed mutagenesis. See Ausubel *et al.*, *supra*, chapter 8, Supplement 37 (1997). In a typical method, a target DNA is cloned into a single-stranded

DNA bacteriophage vehicle. Single-stranded DNA is isolated and hybridized with a oligonucleotide containing the desired nucleotide alteration(s). The complementary strand is synthesized and the double stranded phage is introduced into a host. Some of the resulting progeny will contain the desired mutant, which can be confirmed using DNA sequencing. In addition, various methods are available that increase the probability that the progeny phage will be the desired mutant. These methods are well known to those in the field and kits are commercially available for generating such mutants.

ISOLATING HOMOLOGS

Methods

By using the sequences disclosed herein as probes or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs. "Homologs" are essentially naturally-occurring variants and include allelic, species-specific and tissue-specific variants.

Region-specific primers or probes derived from the nucleotide sequence(s) provided can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog using known methods (Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)). Such an application is useful in diagnostic methods, as described in more detail below, as well as in preparing full-length DNAs from various sources. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. As a general guide, the formula $3(G+C) + 2(A+T) = ^\circ\text{C}$, is useful.

When using primers derived from the inventive sequences, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C), only sequences with greater than 75% sequence identity to the primer will be amplified. By employing lower stringency conditions (*e.g.*, annealing at 35-37°C), sequences which have greater than 40-50% sequence identity to the primer also will be amplified.

The PCR product may be subcloned and sequenced to confirm that it indeed displays the expected sequence identity. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled

and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see e.g., Sambrook et al., 1989, *supra*.

When using DNA probes derived from the inventive sequences for colony/plaque hybridization, one skilled in the art will recognize that by employing medium to high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPEC and 50% formamide, and washing at 50-65°C in 0.5X SSPEC), sequences having regions with greater than 90% sequence identity to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPEC and 40-45% formamide, and washing at 42°C in SSPEC), sequences having regions with greater than 35-45% sequence identity to the probe will be obtained.

Suitably, genomic or cDNA libraries can be constructed and screened in accord with the previous paragraph. The libraries should be derived from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. The clone containing the homolog may then be purified through methods routinely practiced in the art, and subjected to sequence analysis.

Additionally, an expression library can be constructed utilizing DNA isolated from or cDNA synthesized from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. In this manner, clones may be induced and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal gene product, as described herein. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, *ANTIBODIES: A LABORATORY MANUAL*, Cold Spring Harbor Press, Cold Spring Harbor Press.)

Human Homologs

Any organism or tissue can be used as the source for homologs of the present invention so long as the organism or tissue naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs is human.

PROTEINS OF THE INVENTION

One class of proteins included within the invention is encoded by the inventive DNA molecules presented. Other proteins according to the invention are those encoded by the DNA variants described above. As noted, these variants are designed with the encoded proteins in mind.

A preferred class of protein fragments includes those fragments which retain any biological activity. These molecules share functional features common the family of proteins, although these characteristics may vary in degree.

According to one aspect of the invention fragments of the inventive proteins are contemplated. Some preferred fragments are those which are capable of eliciting an immune response. Generally these "antigenic" fragments will be from about five amino acids in length to about fifty amino acids in length. Some preferred antigenic fragments are from five to about twenty amino acids long. "Antigenic" response may refer to a T cell response, a B cell response or a response by cells of the macrophage/monocyte lineages. In most cases, however, it will refer to the immune response involved in the generation of antibodies. In other words, the relevant immune response is that of helper T cells and/or B cells. These preferred molecules comprise one or more T cell and /or B cell epitopes.

ANTIBODIES OF THE INVENTION

Antibodies raised against the proteins and protein fragments of the invention also are contemplated by the invention. Described below are antibody products and methods for producing antibodies capable of specifically recognizing one or more epitopes of the presently described proteins and their derivatives.

Antibodies include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies including single chain Fv (scFv) fragments, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, epitope-binding fragments, and humanized forms of any of the above.

As known to one in the art, these antibodies may be used, for example, in the detection of a target protein in a biological sample. They also may be utilized as part of treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels or for the presence of abnormal forms of the such proteins.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol. Methods* 35:1-21 (1980); Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983); Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96). Antibodies may also be generated by the known techniques of phage display and *in vitro* immunization.

Polyclonal Antibodies

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as an inventive protein or an antigenic derivative thereof.

Polyclonal antiserum, containing antibodies to heterogeneous epitopes of a single protein, can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified, as known in the art, to enhance immunogenicity. Immunization methods include subcutaneous or intraperitoneal injection of the polypeptide.

Effective polyclonal antibody production is affected by many factors related both to the antigen and to the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and/or adjuvant. In addition, host animal response may vary with site of inoculation. Both inadequate or excessive doses of antigen may result in low titer antisera. In general, however, small doses (high ng to low μ g levels) of antigen administered at multiple intradermal sites appears to be most reliable. Host animals may include but are not limited to rabbits, mice, chickens and rats, to name but a few. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

The protein immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin β -galactosidase) or through the inclusion of an adjuvant during immunization. Adjuvants include Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

Booster injections can be given at regular intervals, with at least one usually being required for optimal antibody production. The antiserum may be harvested when the antibody titer begins to fall. Titer may be determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen. See, for example, Ouchterlony *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μ M). The antiserum may be purified by affinity chromatography using the immobilized immunogen carried on a solid support. Such methods of affinity chromatography are well known in the art.

Affinity of the antisera for the antigen may be determined by preparing competitive binding curves, as described, for example, by Fisher, Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D.C. (1980).

In addition to using protein as the immunogen, DNA molecules may be used directly. In this manner, a DNA encoding the protein immunogen is administered. Boosting and harvesting is done in a manner analogous to that detailed above. Yet another method of producing antibodies entails immunizing chickens and harvesting the antibodies from their eggs.

Monoclonal Antibodies

Monoclonal antibodies (MAbs), are homogeneous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture or *in vivo*. MAbs may be produced

by making hybridomas which are immortalized cells capable of secreting a specific monoclonal antibody.

Monoclonal antibodies to any of the proteins, peptides and epitopes thereof described herein can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495-497 (1975) (and U.S. Patent No. 4,376,110) or modifications of the methods thereof, such as the human B-cell hybridoma technique (Kosbor *et al.*, 1983, *Immunology Today* 4:72; Cole *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80: 2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96).

In one method a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen are isolated.

The spleen cells are fused, typically using polyethylene glycol, with mouse myeloma cells, such as SP2/0-Ag14 myeloma cells. The excess, unfused cells are destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted, and aliquots are plated to microliter plates where growth is continued.

Antibody-producing clones (hybridomas) are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures. These include ELISA, as originally described by Engvall, *Meth. Enzymol.* 70:419 (1980), western blot analysis, radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)) and modified methods thereof.

Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* *BASIC METHODS IN MOLECULAR BIOLOGY*, Elsevier, New York. Section 21-2 (1989). The hybridoma clones may be cultivated *in vitro* or *in vivo*, for instance as ascites. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production. Alternatively, hybridoma culture in hollow fiber bioreactors provides a continuous high yield source of monoclonal antibodies.

The antibody class and subclass may be determined using procedures known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

MAbs may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. Methods of purifying monoclonal antibodies are well known in the art.

Antibody Derivatives and Fragments

Fragments or derivatives of antibodies include any portion of the antibody which is capable of binding the target antigen, or a specific portion thereof. Antibody derivatives include poly-specific (*e.g.*, bi-specific) antibodies, which contain binding sites specific for two or more different epitopes. These epitopes may be from the same or different inventive molecules or one or more epitope may be from a molecule not specifically disclosed here.

Antibody fragments specifically include $F(ab')_2$, Fab, Fab' and Fv fragments. These can be generated from any class of antibody, but typically are made from IgG or IgM. They may be made by conventional recombinant DNA techniques or, using the classical method, by proteolytic digestion with papain or pepsin. See CURRENT PROTOCOLS IN IMMUNOLOGY, chapter 2, Coligan *et al.*, eds., (John Wiley & Sons 1991-92).

$F(ab')_2$ fragments are typically about 110 kDa (IgG) or about 150 kDa (IgM) and contain two antigen-binding regions, joined at the hinge by disulfide bond(s). Virtually all, if not all, of the Fc is absent in these fragments. Fab' fragments are typically about 55 kDa (IgG) or about 75 kDa (IgM) and can be formed, for example, by reducing the disulfide bond(s) of an $F(ab')_2$ fragment. The resulting free sulfhydryl group(s) may be used to conveniently conjugate Fab' fragments to other molecules, such as detection reagents (*e.g.*, enzymes).

Fab fragments are monovalent and usually are about 50 kDa (from any source). Fab fragments include the light (L) and heavy (H) chain, variable (V_L and V_H , respectively) and constant (C_L , C_H , respectively) regions of the antigen-binding portion of the antibody. The H and L portions are linked by an intramolecular disulfide bridge.

Fv fragments are typically about 25 kDa (regardless of source) and contain the variable regions of both the light and heavy chains (V_L and V_H , respectively). Usually, the V_L and V_H chains are held together only by non-covalent interactions and, thus, they readily dissociate. They do, however, have the advantage of small size and they retain the same binding properties of the larger Fab fragments. Accordingly, methods have been developed to crosslink the V_L and V_H chains, using, for example, glutaraldehyde (or other chemical crosslinkers), intermolecular disulfide bonds (by incorporation of cysteines) and peptide linkers. The resulting Fv is now a single chain (*i.e.*, SCFv).

Other antibody derivatives include single chain antibodies (U.S. Patent 4,946,778; Bird, *Science* 242:423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85:5879-5883 (1988); and Ward *et al.*, *Nature* 334:544-546 (1989)). Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain FV (SCFv).

One preferred method involves the generation of scFvs by recombinant methods, which allows the generation of Fvs with new specificities by mixing and matching variable chains from different antibody sources. In a typical method, a recombinant vector would be provided which comprises the appropriate regulatory elements driving expression of a cassette region. The cassette region would contain a DNA encoding a peptide linker, with convenient sites at both the 5' and 3' ends of the linker for generating fusion proteins. The DNA encoding a variable region(s) of interest may be cloned in the vector to form fusion proteins with the linker, thus generating an scFv.

In an exemplary alternative approach, DNAs encoding two Fvs may be ligated to the DNA encoding the linker, and the resulting tripartite fusion may be ligated directly into a conventional expression vector. The scFv DNAs generated any of these methods may be expressed in prokaryotic or eukaryotic cells, depending on the vector chosen.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, *Science*, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Derivatives also include "chimeric antibodies" (Morrison *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Neuberger *et al.*, *Nature*, 312:604-608 (1984); Takeda *et al.*, *Nature*, 314:452-454 (1985)). These chimeras are made by splicing the DNA encoding a mouse antibody molecule of appropriate specificity with, for instance, DNA encoding a human antibody molecule of appropriate specificity. Thus, a chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. These are also known sometimes as "humanized" antibodies and they offer the added

advantage of at least partial shielding from the human immune system. They are, therefore, particularly useful in therapeutic *in vivo* applications.

Labeled Antibodies

The present invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example see (Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer *et al.*, *Meth. Enzym.* 62:308 (1979); Engval *et al.*, *Immunol.* 109:129 (1972); Goding, *J. Immunol. Meth.* 13:215 (1976)). The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* diagnostic assays.

Immobilized Antibodies

The foregoing antibodies also may be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "*Handbook of Experimental Immunology*" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby *et al.*, *Meth. Enzym.* 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immunoaffinity purification of the proteins of the present invention.

THERAPEUTIC AND DIAGNOSTIC COMPOSITIONS

The proteins, antibodies and polynucleotides of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in *Remington's Pharmaceutical Sciences* (16th ed., Osol, A., Ed., Mack, Easton PA (1980)). In order to form a pharmaceutically acceptable composition suitable for effective administration,

such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvate may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-*p*-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the composition may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.* gelatin for

use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

RECOMBINANT CONSTRUCTS AND EXPRESSION

The present invention further provides recombinant DNA constructs comprising one or more of the nucleotide sequences of the present invention. The recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a DNA or DNA fragment, typically bearing an open reading frame, is inserted, in either orientation.

The gene products encoded by the subject DNAs may be produced by recombinant DNA technology using techniques well known in the art. See, for example, the techniques described in Sambrook et al., 1989, *supra*, and Ausubel et al., 1989, *supra*. Alternatively, the DNA sequences may be chemically synthesized using, for example, synthesizers. See, for

example, the techniques described in OLIGONUCLEOTIDE SYNTHESIS, 1984, Gait, ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety. They may be assembled from fragments and short oligonucleotide linkers, or from a series of oligonucleotides. They are preferably made by RT-PCR methods. The resulting synthetic gene is capable of being expressed in a recombinant vector.

In some cases the recombinant constructs will be expression vectors, which are capable of expressing the RNA and/or protein products of the encoded DNA(s). Thus, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the open reading frame (ORF). The vector may further comprise a selectable marker sequence.

Specific initiation signals may also be required for efficient translation of inserted target gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where a target DNA includes its own initiation codon and adjacent sequences is inserted into the appropriate expression vector, no additional translation control signals may be needed. However, in cases where only a portion of an ORF is used, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire target. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner *et al.*, *Methods in Enzymol.* 153:516-544 (1987)). Some appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U.S. Patent No. 5,082,767.

The present invention further provides host cells containing at least one of the DNAs of the present invention. The host cell can be virtually any cell for which expression vectors are available. It may be, for example, a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic

cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis *et al.*, *Basic Methods in Molecular Biology* (1986)).

A wide variety of expression systems are available, such as: yeast (*e.g.* *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the target DNA; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the target DNA sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.* Ti plasmid) containing target DNA coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

Depending on the system chosen, the resulting product may differ. For example, proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

Vectors

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting selection of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequence, and in one aspect of the invention, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal or C-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Bacterial Expression

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, if desirable, to provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may, also be employed as a matter of choice.

Bacterial vectors may be, for example, bacteriophage-, plasmid- or cosmid-based. These vectors can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids typically containing elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, GEM 1 (Promega Biotec, Madison, WI, USA), pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pKK232-8, pDR540, and pRIT5 (Pharmacia).

These "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Bacterial promoters include lac, T3, T7, lambda P_R or P_L, trp, and ara.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed/induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the coding sequence may be ligated into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye et al. 1985, *Nucleic Acids*

Res. 13:3101-3109; Van Heeke *et al.*, 1989, *J. Biol. Chem.* 264:5503-5509); pET vectors, Studier *et al.*, *Methods in Enzymology* 185: 60-89 (Academic Press 1990); and the like.

Moreover, pGEX vectors may be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and easily can be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a one embodiment, full length cDNA sequences are appended with in-frame *Bam*HI sites at the amino terminus and *Eco*RI sites at the carboxyl terminus using standard PCR methodologies (Innis *et al.*, 1990, *supra*) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.* 1985, *EMBO J.* 4: 1075; Zabeau and Stanley, 1982, *EMBO J.* 1:1217).

Eukaryotic Expression

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Mammalian promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Exemplary mammalian vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Selectable markers include CAT (chloramphenicol transferase).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the coding sequence of interest

may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a target protein in infected hosts. (*E.g.*, See Logan *et al.*, 1984, *Proc. Natl. Acad. Sci. USA* 81:3655-3659).

In one embodiment, cDNA sequences encoding the full-length open reading frames are ligated into pCMVB replacing the β -galactosidase gene such that cDNA expression is driven by the CMV promoter (Alam, 1990, *Anal. Biochem.* 188: 245-254; MacGregor *et al.*, 1989, *Nucl. Acids Res.* 17: 2365; Norton *et al.* 1985, *Mol. Cell. Biol.* 5: 281).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins.

Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins in eukaryotic cells, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker.

Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the target protein. Such engineered cell lines may be

particularly useful in screening and evaluation of compounds that affect the endogenous activity of the protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, *Cell* 11:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase (Szybalska *et al.*, *Proc. Natl. Acad. Sci. USA* 48:2026 (1962)), and adenine phosphoribosyltransferase (Lowy, *et al.*, *Cell* 22:817 (1980)) genes can be employed in tk⁻, hgp^rt⁻ or apr^t⁻ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, *et al.*, *Proc. Natl. Acad. Sci. USA* 77:3567 (1980)); O'Hare, *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, *J. Mol. Biol.* 150:1); and hydro, which confers resistance to hygromycin (Santerre, *et al.*, 1984, *Gene* 30:147) genes.

An alternative fusion protein system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 8972-8976 (1991)). In this system, the gene of interest is subcloned into a vaccinia-based plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni²⁺ nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The target coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a target gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (E.g., see Smith *et al.*, 1983, *J. Virol.* 46: 584; Smith, U.S. Patent No. 4,215,051).

While the present proteins can be expressed in recombinant systems, as described above, cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Purification of Recombinant Proteins

Recombinant proteins produced may be isolated by host cell lysis. This may be followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, like lysozyme and chelators.

If inclusion bodies are formed in bacterial systems, they may be extracted from cell pellets using, for example, detergents, reducing agents, salts, urea, guanidinium chloride and extremes of pH (*e.g.* <4 or >10). If denaturation occurs, protein refolding steps (*e.g.*, dialysis) can be used, as necessary, in completing configuration of the mature protein. If disulfide bridges are present in the native protein, they may be reoxidized using known methods.

By way of specific non-limiting example, the recombinant bacterial cells, for example *E. coli*, are grown in any of a number of suitable media, for example LB, and the expression of the recombinant protein induced by adding IPTG (*e.g.*, *lac* operator-promoter) to the media or switching incubation to a higher temperature (*e.g.*, λ cI⁸⁵⁷). After culturing the bacteria for a further period of between 2 and 24 hours, the cells are collected by centrifugation and washed to remove residual media. The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the cell membranes from the soluble cell components. If the protein aggregates into inclusion bodies, this centrifugation can be performed under conditions whereby the dense inclusion bodies are selectively enriched by incorporation of sugars such as sucrose into the buffer and centrifugation at a selective speed. The inclusion bodies can then be washed in any of several solutions to remove some of the contaminating host proteins, then solubilized in solutions containing high concentrations of urea (*e.g.* 8M) or chaotropic agents such as guanidinium hydrochloride in the presence of reducing agents such as β -mercaptoethanol or DTT (dithiothreitol).

At this stage it may be advantageous to incubate the protein for several hours under conditions suitable for the protein to undergo a refolding process into a conformation which

more closely resembles that of the native protein. Such conditions generally include low protein concentrations less than 500 µg/ml), low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulphide bonds within the protein molecule. The refolding process can be monitored, for example, by SDS-PAGE or with antibodies which are specific for the native molecule. Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

Labeling Proteins

When used as a component in assay systems such as those described, below, the target protein may be labeled, either directly or indirectly, to facilitate detection of the present *res*-like molecules either *in vitro* or *in vivo*. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ¹²⁵I; enzyme labeling systems that generate a detectable colorimetric signal or light when exposed to substrate; and fluorescent labels.

Where recombinant DNA technology is used for protein production the, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection. These fusion proteins may, for example, add amino acids which facilitate further chemical modification. They also may add a functional moiety, such as an enzyme, which directly facilitates detection.

TRANSGENIC ANIMALS

The invention further contemplates animal models for studying the function of the present molecules and for overproducing the protein products. The disclosed DNA sequences may be used in conjunction with techniques for producing transgenic animals that are well known to those of skill in the art.

To prepare transgenic animals, target gene sequences may for example be introduced into, and overexpressed in, the genome of the animal of interest, or, if endogenous target gene sequences are present, they may either be overexpressed or, alternatively, be disrupted in order to underexpress or inactivate target gene expression, such as described for the disruption of apoE in mice (Plum *et al.*, *Cell* 71: 343-353 (1992)).

In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the animal and cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation.

For underexpression of an endogenous target gene sequence, such a sequence may be isolated and engineered such that when reintroduced into the genome of the animal of interest, the endogenous target gene alleles will be inactivated. Preferably, the engineered target gene sequence is introduced via gene targeting such that the endogenous target sequence is disrupted upon integration of the engineered target gene sequence into the animal's genome.

Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate cardiovascular disease animal models. Goats, cows and sheep are particularly preferred for producing protein *in vivo*.

Any technique known in the art may be used to introduce a target gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe *et al.*, U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82:6148-6152 (1985)); gene targeting in embryonic stem cells (Thompson *et al.*, *Cell* 56:313-321 (1989)); electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803-1814 (1983)); and sperm-mediated gene transfer (Lavitrano *et al.*, *Cell* 57:717-723 (1989)); *etc.* For a review of such techniques, see Gordon, Transgenic Animals, *Intl. Rev. Cytol.* 115:171-229 (1989).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching

of Lasko et al. (Lasko *et al.*, *Proc. Natl. Acad. Sci. USA* 89:3232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the target gene be integrated into the chromosomal site of the endogenous target gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous target gene of interest are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous target gene.

The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene of interest in only that cell type, by following, for example, the teaching of Gu *et al.* *Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant target gene and protein may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and RT-PCR. Samples of target gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the target gene transgene gene product of interest.

The transgenic animals that express target gene mRNA or target gene transgene peptide (detected immunocytochemically, using antibodies directed against the target gene product's epitopes) at easily detectable levels should then be further evaluated to identify those animals which display characteristic increased susceptibility to carcinogenesis. Additionally, specific cell types within the transgenic animals may be analyzed and assayed *in vitro* for cellular phenotypes characteristic of mutant phenotype.

Once target gene transgenic founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include but are not limited to: outbreeding of founder animals with more

than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound target gene transgenics that express the target gene transgene of interest at higher levels because of the effects of additive expression of each target gene transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order both to augment expression and eliminate the possible need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; breeding animals to different inbred genetic backgrounds so as to examine effects of modifying alleles on expression of the target gene transgene and the possible development of carcinogenesis. One such approach is to cross the target gene transgenic founder animals with a wild type strain to produce an F1 generation that exhibits increased susceptibility to carcinogenesis. The F1 generation may then be inbred in order to develop a homozygous line, if it is found that homozygous target gene transgenic animals are viable.

Methods of generating "knockout" mice using homologous recombination in embryonic stem cells are well known in the art. Suitable methods are described, for example, in Mansour *et al.*, *Nature*, 336:348 (1988); Zijlstra *et al.*, *Nature*, 342:435 (1989) and 344:742 (1990); and Hasty *et al.*, *Nature*, 350:243 (1991). This genomic DNA can be obtained by conventional methods using the cDNA sequence as a probe in a commercially-available genomic DNA library.

Briefly, a genomic fragment is cleaved with a restriction endonuclease and a heterologous cassette containing a neomycin-resistance gene is inserted at the cleavage site. A suitable cassette is the GTI-II *neo* cassette described by Lufkin *et al.*, *Cell* 66:1105 (1991). The modified genomic fragment is cloned into a suitable targeting vector that is introduced into murine embryonic stem cells by electroporation. Cells that have undergone homologous recombination (and hence disruption of the gene) are selected by resistance to G418, and used to generate chimeric mice using well known methods. See Lufkin *et al.*, *supra*. Traditional breeding methods then can be used to generate mice that are homozygous for the disrupted gene.

The phenotype of mice that are homozygous for the mutation then can be studied to provide insights into the role of the protein in, for example, carcinogenesis. These mice also can be used as models for developing new treatments for cancers. If this mutation is lethal in

homozygous mice (for example during embryogenesis) heterozygous mice, which express only half the amount of the protein can also be studied.

GENE THERAPY APPLICATIONS

When mutations in the inventive protein, or in the elements controlling expression of that protein, are found to be associated with a malignant phenotype, control of cellular proliferation can be restored by gene therapy methods. For example, overexpression of the protein can be counteracted by concurrent expression of an antisense molecule that binds to and inhibits expression of the mRNA encoding the protein. Alternatively, overexpression can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. In another embodiment, where expression of a mutated protein induces the malignant phenotype, concomitant expression of the non-mutated molecule via introduction of an exogenous gene may be used. Methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art.

Each of these methods requires a system for introducing a vector into the cells containing the mutated gene. The vector encodes either an antisense or ribozyme transcript of the inventive protein. The construction of a suitable vector can be achieved by any of the methods well-known in the art for the insertion of exogenous DNA into a vector. *See, e.g., Sambrook et al., Molecular Cloning* (Cold Spring Harbor Press 2d ed. 1989), which is incorporated herein by reference. In addition, the prior art teaches various methods of introducing exogenous genes into cells *in vivo*. *See Rosenberg et al., Science* 242:1575-1578 (1988) and Wolff *et al., PNAS* 86:9011-9014 (1989), which are incorporated herein by reference. The routes of delivery include systemic administration and administration *in situ*. Well-known techniques include systemic administration with cationic liposomes, and administration *in situ* with viral vectors. Any one of the gene delivery methodologies described in the prior art is suitable for the introduction of a recombinant vector containing an inventive gene according to the invention into a MTX-resistant, transport-deficient cancer cell. A listing of present-day vectors suitable for the purpose of this invention is set forth in Hodgson, *Bio/Technology* 13: 222 (1995), which is incorporated by reference.

For example, liposome-mediated gene transfer is a suitable method for the introduction of a recombinant vector containing an inventive gene according to the invention

into a MTX-resistant, transport-deficient cancer cell. The use of a cationic liposome, such as DC-Chol/DOPE liposome, has been widely documented as an appropriate vehicle to deliver DNA to a wide range of tissues through intravenous injection of DNA/cationic liposome complexes. See Caplen *et al.*, *Nature Med.* 1:39-46 (1995) and Zhu *et al.*, *Science* 261:209-211 (1993), which are herein incorporated by reference. Liposomes transfer genes to the target cells by fusing with the plasma membrane. The entry process is relatively efficient, but once inside the cell, the liposome-DNA complex has no inherent mechanism to deliver the DNA to the nucleus. As such, the most of the lipid and DNA gets shunted to cytoplasmic waste systems and destroyed. The obvious advantage of liposomes as a gene therapy vector is that liposomes contain no proteins, which thus minimizes the potential of host immune responses.

As another example, viral vector-mediated gene transfer is also a suitable method for the introduction of the vector into a target cell. Appropriate viral vectors include adenovirus vectors and adeno-associated virus vectors, retrovirus vectors and herpesvirus vectors.

Adenoviruses are linear, double stranded DNA viruses complexed with core proteins and surrounded by capsid proteins. The common serotypes 2 and 5, which are not associated with any human malignancies, are typically the base vectors. By deleting parts of the virus genome and inserting the desired gene under the control of a constitutive viral promoter, the virus becomes a replication deficient vector capable of transferring the exogenous DNA to differentiated, non-proliferating cells. To enter cells, the adenovirus fibre interacts with specific receptors on the cell surface, and the adenovirus surface proteins interact with the cell surface integrins. The virus penton-cell integrin interaction provides the signal that brings the exogenous gene-containing virus into a cytoplasmic endosome. The adenovirus breaks out of the endosome and moves to the nucleus, the viral capsid falls apart, and the exogenous DNA enters the cell nucleus where it functions, in an epichromosomal fashion, to express the exogenous gene. Detailed discussions of the use of adenoviral vectors for gene therapy can be found in Berkner, *Biotechniques* 6:616-629 (1988) and Trapnell, *Advanced Drug Delivery Rev.* 12:185-199 (1993), which are herein incorporated by reference. Adenovirus-derived vectors, particularly non-replicative adenovirus vectors, are characterized by their ability to accommodate exogenous DNA of 7.5 kB, relative stability, wide host range, low pathogenicity in man, and high titers (10^4 to 10^5 plaque forming units per cell). See Stratford-Perricaudet *et al.*, *PNAS* 89:2581 (1992).

Adeno-associated virus (AAV) vectors also can be used for the present invention. AAV is a linear single-stranded DNA parvovirus that is endogenous to many mammalian species. AAV has a broad host range despite the limitation that AAV is a defective parvovirus which is dependent totally on either adenovirus or herpesvirus for its reproduction *in vivo*. The use of AAV as a vector for the introduction into target cells of exogenous DNA is well-known in the art. *See, e.g., Lebkowski et al., Mole. & Cell. Biol.* 8:3988 (1988), which is incorporated herein by reference. In these vectors, the capsid gene of AAV is replaced by a desired DNA fragment, and transcomplementation of the deleted capsid function is used to create a recombinant virus stock. Upon infection the recombinant virus uncoats in the nucleus and integrates into the host genome.

Another suitable virus-based gene delivery mechanism is retroviral vector-mediated gene transfer. In general, retroviral vectors are well-known in the art. *See Breakfield et al., Mole. Neuro. Biol.* 1:339 (1987) and Shih *et al.*, in *Vaccines* 85: 177 (Cold Spring Harbor Press 1985). A variety of retroviral vectors and retroviral vector-producing cell lines can be used for the present invention. Appropriate retroviral vectors include Moloney Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus. These vectors include replication-competent and replication-defective retroviral vectors. In addition, amphotropic and xenotropic retroviral vectors can be used. In carrying out the invention, retroviral vectors can be introduced to a tumor directly or in the form of free retroviral vector producing-cell lines. Suitable producer cells include fibroblasts, neurons, glial cells, keratinocytes, hepatocytes, connective tissue cells, ependymal cells, chromaffin cells. *See Wolff et al., PNAS* 84:3344 (1989).

Retroviral vectors generally are constructed such that the majority of its structural genes are deleted or replaced by exogenous DNA of interest, and such that the likelihood is reduced that viral proteins will be expressed. *See Bender et al., J. Virol.* 61:1639 (1987) and Armento *et al., J. Virol.* 61:1647 (1987), which are herein incorporated by reference. To facilitate expression of the antisense or ribozyme molecule, of the inventive protein, a retroviral vector employed in the present invention must integrate into the genome of the host cell genome, an event which occurs only in mitotically active cells. The necessity for host cell replication effectively limits retroviral gene expression to tumor cells, which are highly

replicative, and to a few normal tissues. The normal tissue cells theoretically most likely to be transduced by a retroviral vector, therefore, are the endothelial cells that line the blood vessels that supply blood to the tumor. In addition, it is also possible that a retroviral vector would integrate into white blood cells both in the tumor or in the blood circulating through the tumor.

The spread of retroviral vector to normal tissues, however, is limited. The local administration to a tumor of a retroviral vector or retroviral vector producing cells will restrict vector propagation to the local region of the tumor, minimizing transduction, integration, expression and subsequent cytotoxic effect on surrounding cells that are mitotically active.

Both replicatively deficient and replicatively competent retroviral vectors can be used in the invention, subject to their respective advantages and disadvantages. For instance, for tumors that have spread regionally, such as lung cancers, the direct injection of cell lines that produce replication-deficient vectors may not deliver the vector to a large enough area to completely eradicate the tumor, since the vector will be released only from the original producer cells and their progeny, and diffusion is limited. Similar constraints apply to the application of replication deficient vectors to tumors that grow slowly, such as human breast cancers which typically have doubling times of 30 days versus the 24 hours common among human gliomas. The much shortened survival-time of the producer cells, probably no more than 7-14 days in the absence of immunosuppression, limits to only a portion of their replicative cycle the exposure of the tumor cells to the retroviral vector.

The use of replication-defective retroviruses for treating tumors requires producer cells and is limited because each replication-defective retrovirus particle can enter only a single cell and cannot productively infect others thereafter. Because these replication-defective retroviruses cannot spread to other tumor cells, they would be unable to completely penetrate a deep, multilayered tumor *in vivo*. See Markert *et al.*, *Neurosurg.* 77: 590 (1992). The injection of replication-competent retroviral vector particles or a cell line that produces a replication-competent retroviral vector virus may prove to be a more effective therapeutic because a replication competent retroviral vector will establish a productive infection that will transduce cells as long as it persists. Moreover, replicatively competent retroviral vectors may follow the tumor as it metastasizes, carried along and propagated by transduced tumor cells. The risks for complications are greater, with replicatively competent vectors, however.

Such vectors may pose a greater risk than replicatively deficient vectors of transducing normal tissues, for instance. The risks of undesired vector propagation for each type of cancer and affected body area can be weighed against the advantages in the situation of replicatively competent versus replicatively deficient retroviral vector to determine an optimum treatment.

Both amphotropic and xenotropic retroviral vectors may be used in the invention. Amphotropic viruses have a very broad host range that includes most or all mammalian cells, as is well known to the art. Xenotropic viruses can infect all mammalian cell cells except mouse cells. Thus, amphotropic and xenotropic retroviruses from many species, including cows, sheep, pigs, dogs, cats, rats, and mice, *inter alia* can be used to provide retroviral vectors in accordance with the invention, provided the vectors can transfer genes into proliferating human cells *in vivo*.

Clinical trials employing retroviral vector therapy treatment of cancer have been approved in the United States. See Culver, *Clin. Chem.* 40: 510 (1994). Retroviral vector-containing cells have been implanted into brain tumors growing in human patients. See Oldfield *et al.*, *Hum. Gene Ther.* 4: 39 (1993). These retroviral vectors carried the HSV-1 thymidine kinase (HSV-tk) gene into the surrounding brain tumor cells, which conferred sensitivity of the tumor cells to the antiviral drug ganciclovir. Some of the limitations of current retroviral based cancer therapy, as described by Oldfield are: (1) the low titer of virus produced, (2) virus spread is limited to the region surrounding the producer cell implant, (3) possible immune response to the producer cell line, (4) possible insertional mutagenesis and transformation of retroviral infected cells, (5) only a single treatment regimen of pro-drug, ganciclovir, is possible because the "suicide" product kills retrovirally infected cells and producer cells and (6) the bystander effect is limited to cells in direct contact with retrovirally transformed cells. See Bi *et al.*, *Human Gene Therapy* 4: 725 (1993).

Yet another suitable virus-based gene delivery mechanism is herpesvirus vector-mediated gene transfer. While much less is known about the use of herpesvirus vectors, replication-competent HSV-1 viral vectors have been described in the context of antitumor therapy. See Martuza *et al.*, *Science* 252: 854 (1991), which is incorporated herein by reference.

DIAGNOSTIC METHODS

The present invention also contemplates, for certain molecules described below, methods for diagnosis of human disease. In particular, patients can be screened for the occurrence of cancers, or likelihood of occurrence of cancers, associated with mutations in the encoded protein. DNA from tumor tissue obtained from patients suffering from cancer can be isolated and the gene encoding the protein can be sequenced. By examining a number of patients in this manner, mutations in the gene that are associated with a malignant cellular phenotype can be identified. In addition, correlation of the nature of the observed mutations with subsequent observed clinical outcomes allows development of prognostic model for the predicted outcome in a particular patient.

Screening for mutations conveniently can be carried out at the DNA level by use of PCR, although the skilled artisan will be aware that many other well known methods are available for the screening. PCR primers can be selected that flank known mutation sites, and the PCR products can be sequenced to detect the occurrence of the mutation. Alternatively, the 3' residue of one PCR primer can be selected to be a match only for the residue found in the unmutated gene. If the gene is mutated, there will be a mismatch at the 3' end of the primer, and primer extension cannot occur, and no PCR product will be obtained. Alternatively, primer mixtures can be used where the 3' residue of one primer is any nucleotide other than the nonmutated residue. Observation of a PCR product then indicates that a mutation has occurred. Other methods of using, for example, oligonucleotide probes to screen for mutations are described, for example, in U.S. Patent No. 4,871,838, which is herein incorporated by reference in its entirety.

Alternatively, antibodies can be generated that selectively bind either mutated or non-mutated protein. The antibodies then can be used to screen tissue samples for occurrence of mutations in a manner analogous to the DNA-based methods described *supra*.

The diagnostic methods described above can be used not only for diagnosis and for prognosis of existing disease, but may also be used to predict the likelihood of the future occurrence of disease. For example, clinically healthy patients can be screened for mutations in the inventive molecule that correlate with later disease onset. Such mutations may be observed in the heterozygous state in healthy individuals. In such cases a single mutation event can effectively disable proper functioning of the gene and induce a transformed or malignant phenotype. This screening also may be carried out prenatally or neonatally.

DNA molecules according to the invention also are well suited for use in so-called "gene chip" diagnostic applications. Such applications have been developed by, *inter alia*, Synteni and Affymetrix. Briefly, all or part of the DNA molecules of the invention can be used either as a probe to screen a polynucleotide array on a "gene chip," or they may be immobilized on the chip itself and used to identify other polynucleotides via hybridization to the surface of the chip. In this manner, for example, related genes can be identified, or expression patterns of the gene in various tissues can be simultaneously studied. Such gene chips have particular application for diagnosis of disease, or in forensic analysis to detect the presence or absence of an analyte. Suitable chip technology is described for example, in Wodicka *et al.*, *Nature Biotechnology*, 15:1359 (1997) which is hereby incorporated by reference in its entirety, and references cited therein.

PROTEIN-PROTEIN INTERACTIONS

Due to their similarity to certain known proteins, it is anticipated that some of the inventive protein molecules will interact with another class of cellular proteins. This is particularly true of those molecule containing leucine zipper motifs.

Any method suitable for detecting protein-protein interactions can be employed for identifying interacting targets. Among the traditional methods which can be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Utilizing procedures such as these allows for the identification of GAP gene products. Once identified, a GAP protein can be used, in conjunction with standard techniques, to identify its corresponding pathway gene. For example, at least a portion of the amino acid sequence of the pathway gene product can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (see, *e.g.*, Creighton, 1983, *PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES*, W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained can be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for pathway gene sequences. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and for screening are well-known. (See *e.g.*, Ausubel, *supra*, and *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, 1990, Innis *et al.*, eds. Academic Press, Inc., New York).

Additionally, methods can be employed which result in the simultaneous identification of interacting target genes. One method which detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration purposes only and not by way of limitation. One version of this system has been described (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 9578-9582 (1991)) and is commercially available from Clontech (Palo Alto, CA).

Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to a known protein, in this case an inventive protein, and the other contains the activator protein's activation domain fused to an unknown protein (a putative GAP, for instance) that is encoded by a cDNA which has been recombined into this plasmid as part of a cDNA library. The plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, *lacZ*) whose regulatory region contains the transcription activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene, the DNA-binding domain hybrid cannot because it does not provide activation function, and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology can be used to screen activation domain libraries for proteins that interact with a known "bait" gene product. By way of example, and not by way of limitation, gene products known to be involved in TH cell subpopulation-related disorders and/or differentiation, maintenance, and/or effector function of the subpopulations can be used as the bait gene products. Total genomic or cDNA sequences are fused to the DNA encoding on activation domain. This library and a plasmid encoding a hybrid of the bait gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, the bait gene can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

The present invention, thus generally described, will be understood more readily by reference to the following examples, which are provided by way of illustration and are not intended to be limiting of the present invention.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

EXAMPLES

EXAMPLE I: cDNA Library Construction

cDNA library plates and clones originated from five cDNA libraries that were constructed by directional cloning. These are available through the Resource Center (<http://www.rzpd.de>) of the German Genome Project. In particular, the hfbr2 (human fetal brain; RZPD number DKFZp564) and hfkd2 (human fetal kidney; DKFZp566) libraries were generated using the Smart kit (Clontech), except that PCR was carried out with primers that contained uracil residues to permit directional cloning without restriction digestion and ligation, and were complementary with the pAMP1 (LifeTechnologies) cloning sites for directional cloning. The htes3 (human testes; DKFZp434), hute1 (human uterus; DKFZp586) and hmcfl (human mammary carcinoma; DKFZp727) libraries are conventional (Gubler, U., Hoffman, B.J., (1983), A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269), size-selected cDNA libraries. They are cloned into pSPORT1 (LifeTechnologies) via a NotI site which is introduced during reverse transcription downstream of the oligo dT primer and a SalI site that is introduced by the ligation of a adapters. The human mammary carcinoma library was constructed from MCF7 cells.

The cDNA sequences of this application were first identified among the sequences comprising various libraries. Technology has advanced considerably since the first cDNA libraries were made. Many small variations in both chemicals and machinery have been instituted over time, and these have improved both the efficiency and safety of the process. Although the cDNAs could be obtained using an older procedure, the procedure presented in this application is exemplary of one currently being used by persons skilled in the art. For the

purpose of providing an exemplary method, the mRNA isolation and cDNA library construction described here is for the MCF-7 library (DKFZp727) from which the clones named DKFZphmcf1_xxyyxx were obtained.

The human cell line MCF-7 was grown in DMEM supplemented with 10% fetal calf serum until confluency. 3×10^8 cells were harvested with a cell scraper in PBS. Cells were lysed in buffer containing 0.5 % NP-40 to leave the nuclei intact. The debris was pelleted by centrifugation at $15\,000 \times g$ for 10 minutes at 4 degrees Celsius. Proteins in the supernatant were degraded in presence of SDS and Proteinase K (30 minutes at 56 degrees Celsius). Precipitation of proteins was done in a Phenol/Chloroform extraction, RNA was precipitated from the aqueous phase with Na-acetate and Ethanol. Polyadenylated messages were isolated using Qiagen Oligotex (QIAGEN, Hilden Germany).

First strand cDNA synthesis was accomplished using an oligo (dT) primer which also contained an NotI restriction site. Second strand synthesis was performed using a combination of DNA polymerase I, *E. coli* ligase and RNase H, followed by the addition of a Sall adaptor to the blunt ended cDNA. The Sall adapted, double-stranded cDNA was then digested with NotI restriction enzyme, and fractionated by size on an agarose gel. DNA of the appropriate size was cut from the gel and cast into a second gel in a 90° angle. After electrophoresis in the second dimension, cDNA of the appropriate size was cut from the gel. The agarose block was broken down with help of gelase. The cDNA was purified with help of two phenol extractions and an ethanol precipitation. The cDNA was ligated into Sall/NotI pre-digested pSport1 vector (LifeTechnologies) and transformed into DH10B bacteria.

The libraries were arrayed into 384-well microtiter plates and spotted on high density nylon membranes for hybridization analysis. Filters and clones are available through the Resource Center. Whole plates were distributed to the sequencing partners of the consortium for systematic sequencing.

EXAMPLE II: Sequencing of cDNA Clones

All clones in the 384-well microtiter plates were sequenced from the 5' end. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on

ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry.

The resulting expressed sequence tag (EST) sequences ("r1 ESTs" = sequenced from 5'-end) were analysed for:

- a) the lack of identical matches with known genes.

For this, the EST-sequence was blasted against the cDNA consortiums own database and after that against public databases and (with BLASTn and BLASTx against EMBL/EMBLNEW and assembled ESTs, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings). ESTs which were identical to known genes in more than 100 bp, with less than 2 mismatches, were excluded from further analysis.

- b) the presence of an open reading frame

Open reading frames (ORFs) were detected with an tool developed by Munich Information Center for Protein Sequences (MIPS) called ORF-map. ORF-map visualises potential start and stop-codons. If an ORF without a stop codon was detected in a r1-EST, the sequence was processed further.

- c) the presence of GC rich sequences

A script developed by MIPS computed the GC-content of the r1-sequence, which should be >40%. Writing similar scripts is within the ordinary skill of one in bioinformatics.

- d) the lack of repeat structures

Repeats such as Alu, Line or CA-repeats were detected by blasting (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings) against a repeat-database compiled by MIPS. If a repeat was present within the r1-sequence, the sequence were not processed further.

Novel clones that met all criteria were identified to the sequencers, who then performed 3'-end sequencing of these clones. The resulting 3' ESTs ("s1 ESTs" = sequenced from 3'-end) were checked for

a) the lack of matches with known genes in public databases, and sequences already generated by us.

This was done by blasting against EMBL/EMBLNEW and assembled EST (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings).

b) the presence of polyadenylation signals.

Again only clones matching the selection criteria were chosen to be sequenced completely by the sequencers. Clones were selected after the following criteria:

A very good ORF had at least one BLASTx match to other proteins. A "good ORF" should extend to the 3' end and be longer than ~40 codons. If the ORF started in the r1 sequence, in front of the potential start codon, there should not exist too many competing start codons in frame with the ORF start codon and the start should match the Kozak consensus ATG. If the EST sequence was too short to decide according to the potential ORF, and there were only a few or no start codons in the sequence the GC content of the Sequence should be greater than 40%. The r1 sequences needed not contain an polyA-tail at the 3' end. In addition, the results of the blasting against the assembled human ESTs could help in questionable cases to decide whether to stop or to continue. A hit against these ESTs was an indication to go further.

Clones passing the above-described screening were sequenced in full. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry. Primer walking (Strauss et al., 1986, Specific-primer-directed DNA sequencing. *Anal Biochem.* 154, 353-360) was the preferred sequencing strategy because of the lower redundancy possible compared to random shotgun (Messing, J., Crea, R., Seeburg, H.P. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9, 32-39) methods. Walking primers were generally designed using software (e.g. Haas, S., Vingron, M., Poustka, A., Wiemann, S. (1998) Primer design in large-scale sequencing. *Nucleic Acids Res.* 26, 3006-3012, Schwager, C., Wiemann, S., Ansorge, W. (1995) GeneSkipper: integrated software environment for DNA sequence assembly and

alignment. HUGO Genome Digest 2, 8-9) that permitted complete automation of this usually time consuming process and helped in the parallel processing of large numbers of clones.

EXAMPLE III: Bioinformatics analysis of full length cDNAs

Each sequence obtained was compared on nucleotide level in a stepwise manner to sequences in EMBL/EMBLNEW, EMBL-EST, EMBL-STS using the BLASTn algorithm. Basic Local Alignment Search Tool (BLAST, Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S. F. et al (1990) J Mol Biol 215:403-10) is used to search for local sequence alignments. BLAST produces alignments of both nucleotide (BLASTn) and amino acid sequences (BLASTp or BLASTx) to determine sequence similarity. BLAST is especially useful in determining exact matches or in identifying homologs, because of the local nature of the alignments. While it is useful for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment BLAST approach is to look threshold or cut off score set by the user. BLAST looks for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output. Parameter settings for the BLAST-operations (BLASTN 2.0a19MP-WashU) described were: EMBL-EMBLNEW: H=0 V=5 B=5 -filter seg; EMBL-EST: H=0 E=1e-10 B=500 V=500 -filter seg; EMBL-STS: H=0 V=5 B=5.

Search against EMBL/EMBLNEW was done to determine whether the cDNAs are already known, and also to find out whether the cDNAs are encoded by genomic sequences already sequenced and published/submitted to these databases.

Search against EMBL-EST was performed to get a first impression how abundant a particular cDNA would be and to get information on tissue specificity (so-called "electronic Northern-Blot", e.g. some of the cDNAs derived of the testis library show only hits to ESTs also derived of testis libraries).

The cDNA-sequences were blasted against EMBL-STS to determine STS-sequence-match to the cDNA, thus providing a mapping information to the new cDNA.

The potential protein-sequences were generated automatically by a script searching for the longest open reading frame (ORF) in each of the three forward frames with a minimum length of 90 codons. Next, the automatically generated ORFs were translated into protein sequences. These protein sequences were searched against the non redundant protein data set of PIR/SwissProt/Trembel/Tremblnew (BLASTP 2.0a19MP-WashU, parameter setting: V=7 B=7 H=0 -filter seg). If the script generated more than one ORF, one ORF was chosen manually by the annotater according to the degree of similarity to known proteins, the location of the ORF in the cDNA, the length, the amino acid composition and the content of Prosite-Motifs.

Additionally there was a BLASTx (BLASTX 2.0a19MP-WashU against non redundant protein database comprising PIR/SWISSPROT/TREMBL/TREMBLNEW; parameter-settings were: matrix/home/data/blast/matrix/aa/BLOSUM62 H=0 V=5 B=5 -filter seg) search to find potential frame shift in the complementary cds of the cDNAs and to identify unspliced or partly spliced cDNAs. The protein sequence was then transferred to the PEDANT system, in order to generate additional information on the new proteins. PEDANT (Protein Extraction, Description, and ANalysis Tool, Frishman, D. & Mewes, H.-W. (1997) PEDANTic genome analysis. Trends in Genetics , 13, 415-416) is a platform developed at the Munich Information Center for Protein Sequences (MIPS, Munich, Germany), which incorporates practically all bioinformatics methods important for the functional and structural characterisation of protein sequences. Computational methods used by PEDANT are:

FASTA

Very sensitive protein sequence database searches with estimates of statistical significance. Pearson W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 183, 63-98.

BLAST2

Very sensitive protein sequence database searches with estimates of statistical significance. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-10.

PREDATOR

High-accuracy secondary structure prediction from single and multiple sequences. Frishman, D. and Argos, P. (1997) 75% accuracy in protein secondary structure prediction. *Proteins*, 27, 329-335. Frishman, D. and Argos, P. (1996) Incorporation of long-distance interactions in a secondary structure prediction algorithm. *Prot. Eng.* 9, 133-142.

STRIDE

Secondary structure assignment from atomic coordinates. Frishman, D. and Argos, P. (1995) Knowledge-based secondary structure assignment. *Proteins* 23, 566-579.

CLUSTALW

Multiple sequence alignment. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.

TMAP

Transmembrane region prediction from multiply aligned sequences. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* 237, 182-192.

ALOM2

Transmembrane region prediction from single sequences. Klein, P., Kanehisa, M., and DeLisi, C. Prediction of protein function from sequence properties: A discriminant analysis of a database. *Biochim. Biophys. Acta* 787, 221-226 (1984). Version 2 by Dr. K. Nakai.

SIGNALP

Signal peptide prediction Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10, 1-6.

SEG

Detection of low complexity regions in protein sequences. Wootton, J.C., Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Computers & Chemistry* 17, 149-163.

COILS

Detection of coiled coils. Lupas, A., M. Van Dyke, and J. Stock, "Predicting Coiled Coils from Protein Sequences." *Science* (1991) 252, 1162-1164.

PROSEARCH

Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M., Smith J.E. (1992) ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. *Biotechniques* 13, 919-921.

BLIMPS

Similarity searches against a database of ungapped blocks. J.C. Wallace and Henikoff S., (1992) PATMAT: a searching and extraction program for sequence, pattern and block queries and databases, *CABIOS* 8, 249-254. Written by Bill Alford.

HMMER

Hidden Markov model software . Sonnhammer E.L.L., Eddy S.R., Durbin R. (1997)
Pfam: A Comprehensive Database of Protein Families Based on Seed Alignments. *Proteins*
28, 405-420.

pl

Perl script that returns the amino acid composition, molecular weight, theoretical pI, and expected extinction coefficient of an amino acid sequence. By Fred Lindberg. The parameter-settings were as follows: known3d: score > 100; BLAST: E-value < 10; SCOP: <= 50 Alignments, E-Value < 0.0001; signalp: Y=0.7; untersucht vom N-Terminus her: 50 aa; funcat: E-value < 0.001; BLOCKS: <= 10 hits; BLIMPS: threshold 1100.0; COILS: threshold 0.95; SEG: threshold 20.0; BLAST in report: E-value < 0.001; PIR-KW, superfamilies, EC-Nummern in report: E-value < 0.00001; known3d in report: score > 120

The results of PEDANT analysis, together with the results of the similarity searches, constitute the basis for the structural and functional annotation of the cDNAs and the encoded proteins, as specified below.

EXAMPLE III: CELLULAR LOCALIZATIONS OF GFP-FUSION PROTEINS

Plasmids of cDNA-GFP fusions were transfected into mammalian tissue culture cells and allowed to express the proteins for up to 48 hours. Live cells were imaged at 24 hours and 48 hours after transfection and the localisations recorded. The chart, below, depicts the apparent final cellular localisations of 107 cDNA-GFP fusions.

In order to minimize the possibility of the GFP interfering with protein function and/or localization, two separate populations of cDNAs were generated encoding N-terminal or C-terminal GFP fusions. Clearly this appears to be a crucial strategy, since overall only 56% of the proteins localised to a specific compartment irrespective of the position of the GFP. In the instances where only one fusion localized, the complementary fusion either gave no expression or a nuclear and cytosolic staining - characteristic for GFP alone expression.

Each cDNA in turn was subjected to bioinformatic analysis. Where possible, the potential subcellular localisations of the expressed proteins were determined. This

information was then compared to the actual localisations determined from expression of the GFP-fusion proteins in mammalian cells.

DKFZphfbr2_16c16

group: Cell structure and motility

DKFZphfbr2_16c16.3 encodes a novel 586 amino acid protein with similarity to the human actin binding protein MAYVEN and Drosophila Kelch.

MAVEN is a novel actin binding protein predominantly expressed in brain. Drosophila kelch is involved in the maintenance of ring canal organization during oogenesis. The amino half of the protein including the BTB domain mediates dimerization, while the amino half might allow cross-linking of ring canal actin filaments, thus organising the inner rim cytoskeleton. The kelch repeat domain is necessary for ring canal localisation and believed to mediate an additional interaction, possibly with actin. The new protein shares the features of both proteins and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins.

The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

similarity to Drosophila kelch

complete cDNA, complete cds, EST hits
on genomic level partly encoded by AC005082 and AC006039

Sequenced by Qiagen

Locus: unknown

Insert length: 3028 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2984

```

1 GGGGGCCCCG GGACGCAGCC CAGTTGGTAG CGTCGCTCCC TGAGCGTTTC
51 TAAGGGGGCC GCCCGGCCCT GTCTTTCGGC AGTGGCCGAG CCACCGCCGC
101 CTGCCGCGCG TTCCAGAGCT GGGCGCTGCA GCTGCACTGC CGATCGCCGT
151 GTTTGGTCGA TAGAATCCCC AGTGTGCCCA GAGAGTGC GA CCCCTCGCCC
201 GGCCCGGCGA GCGCCGGGCG TGAACCGAGC TGAGGGAGGA TGGCAGCCTC
251 TGGGGTGGAG AAGAGCAGCA AGAAGAAGAC CGAGAAGAAA CTTGCTGCTC
301 GGAAGAAGC TAAATTGTTG GCGGGTTTCA TGGCGTCAT GAATAACATG
351 CGAACAACAG AAACGTTGTG TGACGTGATC CTCATGGTCC AGGAAAGAAA
401 GATACCTGCT CATCGTGTG TTCTTGCTGC AGCCAGTCAT TTTTAACT
451 TAATGTTTAC AACTAACATG CTTGAATCAA AGTCCTTTGA AGTAGAATC
501 AAAGATGCTG AACCTGATAT TATTGAACAA CTGGTGAAT TTGCTTATAC
551 TGCTAGAATT TCCGTGAATA GCAACAATGT TCAGTCTTTG TTGGATGCAG
601 CAAACCAATA TCAGATTGAA CCTGTGAAGA AAATGTGTGT TGATTTTTTG
651 AAAGAACAAG TTGATGCTTC AAATTGTCTT GGTATAAGTG TGCTAGCGGA
701 GTGTCTAGAT TGCTCTGAAT TGAAGCAAC TGACAGATGAC TTTATTCATC
751 AGCACTTTAC TGAAGTTTAC AAAACTGATG AATTCTTCA ACTTGATGTC
801 AAGCGAGTAA CACATCTTCT CAACCAAGGAC ACTCTGACTG TGAGAGCAGA
851 GGATCAGGTT TATGATGCTG CAGTCAGGTG GTTGAAATAC GATGAGCCTA
901 ATCGCCAGCC ATTTATGGTT GATATCCTTG CTAAAGTCAG GTTTCCTCTT
951 ATATCAAAGA ATTTCTTAAG TAAAACGGTA CAAGCTGAAC CACTTATTCA
1001 AGACAATCCT GAATGCCTTA AGATGGTGAT AAGTGAATG AGGTACCATC
1051 TACTGTCTCC AGAGGACCGA GAAGAACTTG TAGATGGCAC AAGACCTAGA
1101 AGAAAGAAAC ATGACTACCG CATAGCCCTA TTTGGAGGCT CTCACCCACA
1151 GTCTTGTA GA TATTTTAACC CAAAGGATTA TAGCTGGACA GACATCCGCT
1201 GCCCCTTTGA AAAACGAAGA GATGCAGCAT GCGTGTGTTG GGACAATGTA
1251 GTATACATTT TGGGAGGCTC TCAGCTTTTC CCAATAAAGC GAATGGACTG
1301 CTATAATGTA GTGAAGGATA GCTGGTATTC GAAACTGGGT CCTCCGACAC
1351 CTCGAGACAG CCTTGCTGCA TGTGCTGCAG AAGGCATAAT TTATACATCT
1401 GGAGGTTTCA AAGTAGGAAA CTCAGCTCTG TATTTATTTG AGTGCTATGA
1451 TACGAGAACT GAAAGCTGGC ACACAAAGCC CAGCATGCTG ACCCAGCGCT
1501 GCAGCCATGG GATGGTGGAA GCCAATGGCC TAATCTATGT TTGTTGGTGA
1551 AGTTTAGGAA ACAATGTTTC AGGGAGAGTG CTTAATTCCT GTGAAGTTTA
1601 TGATCCTGCC ACAGAAACAT GGACTGAGCT GTGTCCAATG ATTGAAGCCA
1651 GGAAGAAATCA TGGGCTGGTA TTTGTAAAAG ACAAGATATT TGCTGTGGGT
1701 GGTCAAGATG GTTTAGGTGG TCTGGACAAT GTGGAATATT ACGATATTAA
1751 GTTGAACGAA TGAAGATGG TCTACCAAT GCCATGGAAG GGTGTAACAG
1801 TGAATGTGC AGCAGTTGGC TCTATAGTTT ATGCTTGGC TGGTTTTTCTCAG

```



```

2301 AGAAGATTGG CTCATCAGTG AAGCGCAGTA TCTTAGCTCT AGATTCTATT
2351 TTCATGCATC ACAGAAGTGC TATACGGTTA GGTCTGTTTG TGCTCAGTCA
2401 AGAACTAAGA AATAGTATGA ATTGTAAGTC AAGATGGGCA ACTCAGATGG
2451 AGCAGCTTAG TCTCACAGTT TGCTTGTCTA TTTATTTTAT TTAGTGCCAA
2501 ATGTATTCCA TTTTAAAGT AAGCCAGAGT GAGTCAAGGC ATATACACAC
2551 TTTCTCACAA AACTTCCTAA ACAGATTGGG GGGTTTAATA TGTCCTCACTC
2601 CTCATGAAAT ATATTCAATC CACTTAAATA TATTCCATCT TTTTAACATA
2651 AAATGTAAG CTTAGCACCC ATCATTAAAT TATGTCTCTG TTTTATCCAG
2701 TGGTTAAAAA AGGATTCTGC CTCCTTAGTC CTCCTGTTA AATAAAACCC
2751 AATCATAGTA AGTGATTAC TAGCAAAAG TAAAGCTATT TATAGCAAAT
2801 TTCTAGATCA TTAGAAAAGC ACTGGTAGTT GTACAATATC AGTGTTGACT
2851 TTGAACCTCT TTAACGAGAT CATGAATTCT TTTCCCTTAG CCAAAACATG
2901 AAATATTAA CCTAGTTGTC TCTAAAAGTT TTGTAATCAT GAGTTAGATA
2951 TATGTCATCT CCTATTCATT GCTTTTATGT GATCAATAAA TCTTTTACAA
3001 ACCCAAAAGA AAAAAAAAAA AAAAAAAA

```

BLAST Results

Entry AC005082 from database EMBL:
Homo sapiens clone RG271G13; HTGS phase 1, 7 unordered pieces.
Score = 6460, P = 0.0e+00, identities = 1292/1292
4 exons matching Bp 1180-3007

Entry AC006039 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0319F03; HTGS phase 1, 3 unordered pieces.
Score = 1780, P = 2.0e-117, identities = 368/377
5 exons matching Bp 6-860

Entry HSG20603 from database EMBL:
human STS A005Y34.
Score = 670, P = 1.0e-23, identities = 134/134

Medline entries

93201592:
kelch encodes a component of intercellular bridges in
Drosophila egg chambers.

97412177:
Drosophila kelch is an oligomeric ring canal actin organizer.

Peptide information for frame 3

ORF from 240 bp to 1997 bp; peptide length: 586
Category: strong similarity to known protein

```

1 MAASGVEKSS KKKTEKKLAA REEAKLLAGF MGVNMNMRKQ KTLCDVILMV
51 QERKIPAHRV VLAASHFFN LMFTTNMLES KSFEVELKDA EPDIIEQLVE
101 FAYTARISVN SNNVQSLLDA ANQYQIEPVK KMCVDFLKEQ VDASNCLGIS
151 VLAECIDCPE LKATADDFIH QHFTEVYKTD EFLQLDVKRV THLLNQDTLT
201 VRAEDQVYDA AVRWLKYDEP NRQPFMVVIL AKVRFPPLISK NFLSKTVQAE
251 PLIQDNPECL KMVISGMRYH LLSPEDREEL VDGTTPRRKK HDYRIALFGG
301 SQPQSCRYFN PKDYSWTDIR CPFEEKRDAA CVFWDNVVYI LGGSQLEPIK
351 RMDCYNVVVD SWYSKLGPPPT PRDSLAAACAA EGKIYTSGGG EVGNSALYLF
401 ECDYDRTESW HTKPSMLTQR CSHGMVEANG LIYVCGGSLG NNVSGRVLNS
451 CEVYDPATET WTELCPMIEA RKNHGLVFVK DKIFAVGGQN GLGGLDNVEY
501 YDIKLNWK M VSPMPWKGVV VKCAAVGSIV YVLAGEFGVG RLGHILEYNT
551 ETDKWVANSK VRAFPVTSCL ICVVDTCGAN EETLET

```

BLASTP hits

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN).
Length = 689
Score = 816 (287.2 bits), Expect = 1.9e-81, P = 1.9e-81
Identities = 187/542 (34%), Positives = 290/542 (53%)

Entry AC004021_1 from database TREMBL:
WUGSC:H_DJ0186K10.1"; Human PAC clone DJ0186K10 from 5q31,
complete sequence. Homo sapiens (human)
Length = 497

Entry A45773 from database PIR:
kelch protein, long form - fruit fly (*Drosophila melanogaster*)
Length = 1476
Score = 817 (287.6 bits), Expect = 1.7e-80, P = 1.7e-80
Identities = 189/549 (34%), Positives = 292/549 (53%)

Report for DKFZphfbr2_16c16.3

```
SEQ      SWYSLGPPPTPRDSLAAACAEKGIYTSGGSEVGNALYLFECDTRTESWHTKPSMLTQR
SEG      .....
PRD      CCCCCCCCCCCCCCeeeeeeccceeeeeecccccccceeeeecccccccccccccccccc
```



```

SEQ  CSHGMVEANGLIYVCGSLGNNVSGRVLNSCEVYDPATETWTELCPMIEARKNHGLVFVK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  DKIFAVGGQNGLGGLDNVEYYDIKLNEWKMVSPMPWKGVTVKCAAVGSIVYVLAGFQGVG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  RLGHILEYNTETDKWVANSKVRAFVPTSCLICVVDTCGANEETLET
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_16c16.3

PS00001	442->446	ASN_GLYCOSYLATION	PDOC00001
PS00004	11->15	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	188->192	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	315->319	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	405->409	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	550->554	CK2_PHOSPHO_SITE	PDOC00006
PS00007	202->209	TYR_PHOSPHO_SITE	PDOC00007
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	389->395	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	436->442	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00008	493->499	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16c16.3)

DKFZphfbr2_16f21

group: brain derived

DKFZphfbr2_16f21 encodes a novel 208 amino acid protein with strong similarity to human zinc finger protein 216.

The novel protein shows strong similarity to the human zinc finger protein 216, but has no Zn finger.

PROSITE: Contains no Zinc finger; No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to zinc finger protein 216

complete cDNA, complete cds, EST hits
start matches Kozak consensus ANNatgG,

Sequenced by Qiagen

Locus: unknown

Insert length: 1512 bp

Poly A stretch at pos. 1490, polyadenylation signal at pos. 1474

```
1 GGGAGCAAGC AGGGGTTTCGG CGGCATTACC TGTACCCATT CACCGGCGGC
51 TACCGGCGGC GGC GCGTAGC GTGTCAGGCG GAGAGACCCG CCGCCAGGTG
101 TGCAACTGAG GAACATGGCT CAAGAACTA ATCAGAGCCA AGTGCCTATG
151 CTTTGTTCCT CTGGCTGTGG ATTTTATGGA AACCTCGTA CAAATGGCAT
201 GTGTTTCAGTA TGCTATAAAG AACATCTTCA AAGACAGAAT AGTAGTAATG
251 GTAGAATAAG CCCACCTGCA ACCTCTGTCA GTAGTCTGTC TGAATCTTTA
301 CCAGTTCAAT GCACAGATGG CAGTGTGCCA GAAGCCAGT CAGCATTAGA
351 CTCTACATCT TCATCTATGC AGCCAGCCCC TGTATCAAA CAGTCACTTT
401 TATCAGAATC TGTAGCATCT TCTCAATTGG ACAGTACATC TGTGGACAAA
451 GCAGTACCTG AACACAGAAGA TGTGCAGGCT TCAGTATCAG ACACAGCACA
501 GCAGCCATCT GAAGAGCAAA GCAAGCCTCT TGAAAAACCG AAACAAAAAA
551 AGAATCGCTG TTTTCATGTC AGGAAGAAAG TGGGACTTAC TGGGTTTGAA
601 TGCCGGTGTG GAAATGTTTA CTGTGGTGTA CACCGTTACT CAGATGTACT
651 CAATTGCTCT TACAATTACA AAGCCGATGC TGCTGAGAAA ATCAGAAAAG
701 AAAATCCAGT AGTTGTTGGT GAAAAGATCC AAAAGATTG AACTCCTGCT
751 GGAATACAAA ATTCTTGAGC ATCTGCAAA CAAAATTGA CTTGAGGTTT
801 TTTTTCCTCT AGTCATTGGG AATGTAGAGC AGTGTATCTT GCATGTCATC
851 GGAAGAATAG ATTTTGTGTT TGGTTTGTG TTGAAAATGA CTCTGAACAT
901 TTATTTCCAT TGCAATTCT GTGGCTGAGG AGACTTAAAC TTTACAAGTA
951 TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT
1001 AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTCCA GTATTAAACA
1051 TGCATGCATT AATCTTGCG TTTATTTTCT CATTATGTAT GTATATATCG
1101 CTTTCTCTCT CAGCAGGATT TCTCTTTTGA TAATGCCCTT TAGGGCACA
1151 CTAGTTATCA GTAACGAAT GTATCTTAAT CATTATGGCT GCTTCTGTTT
1201 TTTTATTAAC AAAGGTTATT CATATGTTAG CATATAGTTT CTTTGCACCC
1251 ACTATTTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT
1301 TGTAAAGTTG TGTGTTTAAA CTTTTTTTGG AGCGAGGGAA GAAAAAGCTG
1351 TATGCATTTT ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT
1401 TTGTGTGTAT ACAATATGAA GAATGATCTG AAGTAATTGT GCTGTATTTA
1451 TGTTTATTCA CCAGTCTTTG ATTAAATAAA AAGGAAAACC AGAAAAAAA
1501 AAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 115 bp to 738 bp; peptide length: 208
Category: strong similarity to known protein

```

1 MAQETNHSQV PMLCSTGCGF YGNPRTNGMC SVCYKEHLQR QNSSNGRISP
51 PATSVSSLSE SLPVQCTDGS VPEAQSALDS TSSSMQSPV SNQSLLESV
101 ASSQLDSTSV DKAVPETEDV QASVSDTAQQ PSEEQSKPLE KPKQKKNRCF
151 MCRKKVGLTG FECRCGNVYC GVHRYSDVLN CSYNYKADAA EKIRKENPVV
201 VGEKIQKI

```

BLASTP hits

Entry ATF7H19_1 from database TREMBLNEW:

gene: "F7H19.10"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSAII project) >TREMBL:ATT12H17_21
gene: "T12H17.210"; product: "predicted protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII project)
Score = 206, P = 2.1e-24, identities = 51/146, positives = 77/146

Entry PVPVPR3A_1 from database TREMBL:

gene: "PVPVPR3"; P.vulgaris PVPVPR3 protein mRNA, complete cds.
Score = 237, P = 4.9e-20, identities = 50/136, positives = 73/136

Entry AF062072_1 from database TREMBL:

gene: "ZNF216"; product: "zinc finger protein 216"; Homo sapiens zinc finger protein 216 (ZNF216) gene, complete cds.
Score = 591, P = 1.6e-57, identities = 124/215, positives = 147/215

Alert BLASTP hits for DKFZphfbr2_16f21, frame 1

TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc finger protein ZNF216 mRNA, complete cds., N = 1, Score = 590, P = 2.1e-57

TREMBLNEW:AB001773_1 gene: "pem-6"; product: "PEM-6"; Ciona savignyi pem-6 (posterior end mark 6) mRNA, complete cds., N = 1, Score = 421, P = 1.7e-39

>TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc finger protein ZNF216 mRNA, complete cds.
Length = 213

HSPs:

Score = 590 (88.5 bits), Expect = 2.1e-57, P = 2.1e-57
Identities = 123/213 (57%), Positives = 146/213 (68%)

```

Query:      1 MAQETNHSQV PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPAT---SVSS 57
             MAQETN + PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQ +S GR+SP T S S
Sbjct:      1 MAQETNQTGPGMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNS-GRMSPMGASGSNSP 59

Query:     58 LSESLPVQCTDGSVPEAQSALDSTSSSMQSPVSNQSLLE--SVASSQLDSTSVDKAVP 115
             S+S VQ D + + A STS + PV+ + + ++ S+ D + K
Sbjct:     60 TSDSASVQRADAGLNNCEGAAGSTSEKSRNPVVAALPVTQQTMSISREDKITTPKT-E 118

Query:    116 ETEDVQASVSDTAQQPSEEQS--KPLEKPKQKKNRCFMCRRKKVGLTGFECCRCGNVYCGVH 173
             +E V S + QPS QS K E PK KKNRCFMCRRKKVGLTGF+CRCGN++CG+H
Sbjct:    119 VSEPVTQPSPSVSQPSQSSQSEKAPLPPKPKKNRCFMCRRKKVGLTGFDCCRCGNLFCGLH 178

Query:    174 RYSDVLNCSYNYKADAAEKIRKENPVVVVGEKIQKI 208
             RYSD NC Y+YKA+AA KIRKENPVVV EKIQ+I
Sbjct:    179 RYSDKHNCYPDYKAEAAKIRKENPVVVVAEKIQRI 213

```

Pedant information for DKFZphfbr2_16f21, frame 1

Report for DKFZphfbr2_16f21.1

```

[LENGTH]      208
[MW]           22541.23
[pI]           6.80
[HOMOL]        TREMBL:AF062072_1 gene: "ZNF216"; product: "zinc finger protein 216"; Homo
                sapiens zinc finger protein 216 (ZNF216) gene, complete cds. 9e-57
[PIRKW]        zinc 8e-13
[PIRKW]        zinc finger 8e-13

```


[PIRKW] fusion protein 8e-13
 [SUPFAM] unassigned ubiquitin-related proteins 8e-13
 [SUPFAM] ubiquitin homology 8e-13
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] ASN_GLYCOSYLATION 4
 [KW] Irregular
 [KW] LOW_COMPLEXITY 7.21 %

SEQ MAQETNHSQVPMCLSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPATSVSSLSE
 SEG
 PRD ccc

SEQ SLPVQCTDGSVPPEAQSALDSTSSSMQPSFVSNQSLSESVASSQLDSTSVDKAVPETEDV
 SEGXXXXXXXXXXXXXXXXX.....
 PRD ccc

SEQ QASVSDTAQQPSEEQSKPLEKPKQKNRCFCMRKKVGLTGFECCRCGNVYCGVHRYSDVLN
 SEG
 PRD ccc

SEQ CSYNYKADAAEKIRKENPVVVGEKIQKI
 SEG
 PRD ccchhhhhhhhhhhhhcccccccccccc

Prosites for DKFZphfbr2_16f21.1

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	42->46	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	180->184	ASN_GLYCOSYLATION	PDOC00001
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16f21.1)

DKFZphfbr2_16g18

group: cell cycle

DKFZphfbr2_16g18.3 encodes a novel 984 amino acid protein with similarity to centromeric proteins of yeasts.

The novel protein shows similarity to *S. pombe* SPAC17A5.07c and the *S. cerevisiae* Smt4p suppressor of MIF2 gene. MIF2 encodes a centromeric protein with homology to the mammalian centromeric protein CENP-C. Mutations in MIF2 stabilise dicentric minichromosomes and confer high instability to chromosomes that bear a cis-acting mutation in element I of the yeast centromeric DNA (CDEI). Therefore the new protein should be involved in centromer organisation, too.

The new protein can find application in modulating/blocking the cell cycle and influencing the behavior of chromosomes, both natural and artificial in eukaryotic cells.

similarity to KIAA0797 and yeast Smt4p

complete cDNA, complete cds, EST hits
the yeast Smt4 protein seems to be involved in centromer function
and microtubule organisation

Sequenced by Qiagen

Locus: unknown

Insert length: 4826 bp

Poly A stretch at pos. 4756, polyadenylation signal at pos. 4736

```

1 GGGTCGAGGT CGACGGTATC GATAAGTTTT TTTTTTTTTT TTTTTTTTTT
51 TTTTCTTTTC CCCTCCCCCT CCCTCTCCAA GCCGGAGGGG TCCTGAGGTG
101 ACAGCGCCTG CAACTGAAAT TTCAGCAGCG GGAGAAGATG GACAAGAGAA
151 AGCTCGGGCG ACGGCCATCT TCATCCGAAA TCATCACAGA AGGAAAAAGG
201 AAAAAGTCAT CTTCTGATTT ATCGGAGATA AGAAAGATGT TAAATGCAAA
251 ACCAGAGGAT GTCCATGTTC AATCACCCT GTCCAAATTC AGAAGCTCAG
301 AACCTGAGAC TCTCCCTTTG CAGTGGGAAA GAAGCCTAAG GAATAAAGTC
351 ATCTCTCTAG ACCATAAAAA TAAAAACAT ATCCGAGGGT GTCCTGTTAC
401 TTCCAGGTCA TCACCAGAAA GGATACCCAG AGTTATATTG ACGAATGTCC
451 TGGGAACGGA GTTAGGAAGA AAATACATAA GGACCCACC TGTAAGTCTGAG
501 GGAAGTTTGA GTGATACAGA CAACTTGCAA TCAGAGCAAC TTTCTTCATC
551 ATCTGATGGC AGCCTAGAAAT CTTATCAAAA TCTAAACCCT CACAAGAGCT
601 GTTATTTATC TGAAAGGGGC TCACAACGAA GTAAGACAGT AGATGACAAT
651 TCTGCAAAAGC AGACTGCGCA CAATAAAGAA AAACGAAGAA AGGATGATGG
701 CATTCTCTTT TTAATATCTG ATACTCAGCC TGAAGACCTT AACAGTGGAA
751 GTAGAGGTTG TGATCATCTC GAACAGGAAA GCAGAAACAA GGATGTTAAA
801 TATTCTGATT CAAAAGTGGG ACTCACTCTG ATTTCCAGGA AGACAAGAG
851 AAGGCTTAGA AATAATTTAC CTGATTCTCA ATATTGTACT TCTTTGGATA
901 AGTCAACAGA ACAGACAAAA AAACAAGAAG ATGACTCAAC AATATCCACT
951 GAGTTTGAAA GGCCAAGTGA AAATATCAT CAGGATCCAA AACTGCCTGA
1001 AGAAATTACA ACTAAACCTA CAAAAGTGA TTTTACTAAG CTATCCTCAC
1051 TTAACAGTCA GGAGTTGACT TTGAGTAATG CCACCAAAG TGCCTTGCC
1101 GGTTCACCCA CTGAAACCGT TGAGTACTCT AATTCCATTG ATATTGTGGG
1151 GATTCTTCC CTGGTTGAGA AGGATGAGAA TGAGTTGAAT ACCATAGAAA
1201 AGCCTATTCT AAGAGGACAT AATGAAGGA ACCAATCACT GATCTCAGCT
1251 GAACCAATTG TTGTTTCCAG TGATGAAGAA GGACCTGTTG AACATAAAG
1301 TTCAGAAATT CTTAAGTTAC AATCTAAGCA AGACCGTGAG ACAACTAATG
1351 AAAATGAGAG TACTTCTGAA TCAGCATTGT TAGAACTACC ATTGATTACA
1401 TGTGAATCTG TACAGATGTC ATCTGAATTA TGCCCATATA ATCCTGTCAT
1451 GGAGAACATT TCCAGTATTA TGCCTAGTAA TGAGATGGAT CTACAAGTGG
1501 ATTTTATATT TACTTCTGTT TATATTGGTA AAATAAAAGG AGCTTCTAAA
1551 GGTGTGTTTA CAATCACAAA AAAATATATT AAGATCCCAT TTCAAGTGTC
1601 CCTGAATGAG ATTTCAATGC TAGTGGATAC CACACATTTA AAGCGGTTTG
1651 GGTATGGGAA AAGTAAGGAT GATAATCACA GTAAAAGGAG TCATGCTATT
1701 CTTTCTTCTT GGGTCTCTTC AGATTATCTT CAAGAGATTC AGACCCCAAT
1751 AGAACACTCT GTATTAAGCC AGCAATCAAA ATCTAGTGAA TTCATTTTCC
1801 TTGAACATCA CAATCCTGTT TCACAGAGAG AAGAATTGAA GCTGAAAGAT
1851 ATTATGACCG AAATAAGTAT AATCAGTGGA GAATTAGAGC TTTCTTACCC
1901 GTTGCTTTGG GTTCAGGCAT TTCCTTTGTT TCAGAACCTC TCTTCAAAG
1951 AAAGTTCTTT TATTCAATTAT TACTGTGTTT CAACTTGTTT TTTCCCTGCT
2001 GGTGTGTGCT TTGCTGAAGA AATGAAGCTG AAATCAGTAT CTCAGCCCTC
2051 AAACACAGAT GCGGCCAAGC CTACTTACAC CTTCTGCAAG AAGCAAAGTA
2101 GCGGTTGCTA CTCCCTTTCT ATTACATCTA ATCCAGATGA AGAATGGCGG
2151 GAAGTCAGGC ACACTGGACT TGTTCAAGAG TTGATTGTAT ATCCTCCACC
2201 ACCTACTAAG GGGGATTTGG GAGTAACTAA TGAAGATCTG GAGTGTTTAG
2251 AAGAAAGGAGA GTTTCTTAAT GATGTAATCA TTGATTTTAA CCTTAAGTAT
2301 CTTATATTGG AGAAGGCATC AGATGAACCT GTTGAACGAA GTCACATTTT

```



```

2351 TAGTAGCTTT TTCTATAAAT GCTTGACAAG AAAGGAAAAT AATTTAACAG
2401 AAGATAATCC AAATCTTTCA ATGGCACAGA GAAGACATAA AAGAGTAAGA
2451 ACATGGACTC GTCACATAAA CATTTTTAAAT AAAGATTACA TCTTTGTACC
2501 TGTAATGAG TCGTCTCACT GGTATCTCGC AGTCATTTGT TTTCCATGGT
2551 TAGAAGAAGC TGTGTATGAA GATTTTCCAC AAACGTATATC CCAGCAGTCC
2601 CAGGCTCAGC AGTCCCAAAG TGACAACAAA ACAATAGATA ATGATCTACG
2651 TACTACTTCG ACACGTCTTT TGAGTGCAGA GGATTCCCAA AGTACCCAGT
2701 CGAATATGTC AGTACCAAAG AAAATGTGTA AAAGGCCATG TATTCTTATA
2751 CTAGACTCCT TGAAAGCTGC TTCTGTACGA AACACAGTTC AGAATTTACG
2801 AGAGTATTTA GAGGTAGAGT GGGAAAGTTAA ACTAAAAACT CATCGTCAAT
2851 TCAGCAAAAC AAACATGGTG GATCTATGCC CTAAAGTTCC TAAACAGGAC
2901 AATAGCAGTG ATTGTGGAGT ATATTTATTG CAGTATGTGG AAAGCTTCTT
2951 CAAGGATCCT ATTGTAACT TTGAACTTCC AATTCATTTG GAGAAGTGGT
3001 TTCTCTGTC TAATAAAG ACCAAACGGG AAGATATTCG AGAGCTCATC
3051 TTGAAACTTC ATTTACAGCA ACAGAAGGGC AGCAGTAGCT AGTTAATCTG
3101 TACAACATG ACACAGATGT TCTCTAAGAT TACTGGAAAG CCCCTTACCA
3151 GCATTTGTGT TAGCCAGCTC ACAGAGAAGA AAATAACTTG CAGTAGTTTT
3201 ATAATAAGTC ATTGGAACAT TATTTAAAAT ATGTAGGACA CATTATTAGA
3251 ATTGTTGGGA TCTCATAGAT GGAATGGGAA TGGGGGTGAT ATAGATAAAC
3301 TTACTAGATA TAAATTAAAA TTTTATAAAT ATTTCATATT TTTCTGAGTA
3351 AATATGATTG GATTATGCAA CAGCATATGT AATATGGGAA TGTTTGTAG
3401 ATAATAAAAC TTACATGATC TGTACTTCCA CGTGACTGGG TGCTGAGGGG
3451 AGTTAAAGCC TCCCTGGTGC CAGCCCCAGT GCTTGTCAA TTTGCTGACA
3501 GGTACATCA TATTGTAATT CTATTCTTTG CAGCTCAAGC ATGCAGTATG
3551 AATACTGTGT ATTTTTTAAA AAAATAATTT AGTATCAAGG CTTCAGAAAA
3601 TGCCATTTAC GGCATCCCTT CTGTATGTAA CAAAAGACA TTCATAATGT
3651 TAGGAAGATG ATAAAAATTC GCTCTTTTAA AGTGCAGCTT ATTATTCTCA
3701 ATTGCTAAAT ACGATTACTC TGCTTTTTTT TTTTCATTTT TTTTGATGTC
3751 ATATGTGAGT ATCTTATAAT TTAGTTCATT TGTTCAGGGT AAAATTTGAA
3801 ACAAAAAAAT TTACCTGTGC AAAATAGTTT TTTAAAAAAT ATACATGTAG
3851 CTCAACTTGA GGTACTGCTA TATAAATATT CACTCACATT ATCACGGAAT
3901 TTATGTATAG TTTCTCTAAT ATAGAAGATA AAATGGGTGT CCTCATAACT
3951 TTAACAAAGA AAACCCCTCAG TCCTATTTAT TAATGGGTAG AATTAATAT
4001 ATAATTTTAT AGCTCAGTTT ACCCAGTATT CATCTGCAA GCCAGATTGC
4051 TCTCATTGCT TTTATATTTT TAAATGTAG CTTTATAGAGA CCTATGATCC
4101 TCATGGAATC TAATTTTTTA TAAATATTC AGGTAACAGT TCTGAATTC
4151 TGTGATAATG GTGGCATTAT ATATGATTAA ACACTTCAGA ACTTTCTAAT
4201 GTTATCAGGA GTATTTTGAG GGAGATATGA TTATATTGTA TTTTCTCAGA
4251 TAAGAAAAAT GTTTTTTAA AATATTATTT TAATCTGTTT TAAGCATCTC
4301 TTAGATTTAC ATTATAACTA CATAAAGCAG TGAAGCAAAG GCAAATTAAG
4351 ATAAAGCTAG AAAGCTGAA CATTTTATTT CAAAATCATA CGAATCGGGG
4401 TCAGTTAAGC CTCAGTATTC TTAGCTTTTG TTGATTTTGG CACTATCTTT
4451 ATATTATTAA ATATATTTGT TGTTTGGATA TTTTATATAA AGATGGCTAT
4501 AATTACATAT TTCATTTCCA ATTTGTGTGT GTTGGGGGGT ACTTTTAAAG
4551 GTGACTATTG TTTTGTACAT CTAATTTTGG GAAACCAAGT CTATAAGACA
4601 TCTTGTGATT TCTTAATGTT TTTGTTTGTG TGTTTTCAA AGATATCACT
4651 GTCTTTTATC ATGTTTGTAA GATTGTTTAA AATTCATTTT CCTAAATTAA
4701 TGTGCAAGTA ATGTTTGTAG GATATCGGTG TTTTATATTA AACATATTTT
4751 CAATTCAAAA AAAAAAATAA AAAAATTAT CGATACCGTC GACCTCGATG
4801 ATGATGATGA TGATGATGAT GTCGAC

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 138 bp to 3089 bp; peptide length: 984
 Category: similarity to known protein

```

1  MDKRLGRRP SSSEIITEGK RKKSSSDLSE IRKMLNAKPE DVHVQSPLSK
51  FRSSERWTLF LOWERSLRNK VISLDHKNKK HIRGCPVTSR SSPERIPRVI
101 LTNVLGTELG RYIRTPPVPT EGSLSDTDNL QSEQLSSSSD GSLESYQNLN
151 PHKSCYLSEF GSQRSKTVDD NSAKQTAHKK EKRRKDDGIS LLISDTQPED
201 LNSGSRGCDH LEQESRNKDV KYSDSKVELT LISRKTRRRL RNNLPDSQYC
251 TSLDKSTEQT KQEDDSTIS TEFERPSSENY HQDPKLPFEEI TTKPTKSDFT
301 KLSSLNSQEL TLSNATKSAS AGSTTETVEY SNSIDIVGIS SLVEKDENEL
351 NTIEKPILRG HNEGNSQLIS AEPIVVSSDE EGPVEHKSSE ILKLQSKQDR
401 ETTNENESTS ESALLEPLI TCESVQMSSE LCPYNPVMEN ISSIMPSNEM
451 DLQLDFITS VYIGKIGAS KGCVTITKKY IKIPFQVSLN EISLLVDTH

```



```

501 LKRFGWLWKS DDNHSKRSHA ILFFWVSSDY LQEIQTQLEH SVLSQQSKSS
551 EFIFLELHNP VSQREELKLLK DIMTEISIIS GELELSYPLS WVQAFPLFQN
601 LSSKESSEFIH YYCVSTCSFP AGVAVAEEMK LKSVSQPSNT DAAKPTYTFL
651 KQSSSGCYSL SITSNPDEEW REVRHTGLVQ KLIVYPPPT KGGLGVTNED
701 LECLEEGEFL NDVIIDFYLYL LILEKASDE LVERSHIFSS FFYKCLTRKE
751 NNLTEDNPNL SMAQRRHKRV RTWTRHINIF NKDYIFVPVN ESSHWYLAVI
801 CFPWLEEAUV EDFFQTVSQ SQAQSQSDN KTIDNDLRTT STLSLSAEDS
851 QSTESNMSVP KKMCKRPCIL ILDSLKAASV RNTVQNLREY LEVEWEVKLK
901 THROFSKTNM VDLCPKVPKQ DNSSDCGVYL LQYVESFFKD PIVNFELPIH
951 LEKWFPRHVI KTKREDIREL ILKLHLQQQK GSSS

```

BLASTP hits

Entry SPAC17A5_7 from database TREMBL:
 "SPAC17A5.07c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c17A5. Schizosaccharomyces pombe (fission
 yeast)
 Length = 652
 Score = 275 (96.8 bits), Expect = 1.9e-29, Sum P(3) = 1.9e-29
 Identities = 56/120 (46%), Positives = 78/120 (65%)

Entry S49947 from database PIR:
 SMT4 protein - yeast (Saccharomyces cerevisiae)
 Length = 1034
 Score = 163 (57.4 bits), Expect = 4.6e-16, Sum P(3) = 4.6e-16
 Identities = 46/159 (28%), Positives = 76/159 (47%)

Entry YQG6_CAEEL from database SWISSPROT:
 HYPOTHETICAL 35.7 KD PROTEIN C41C4.6 IN CHROMOSOME II.
 Length = 342
 Score = 162 (57.0 bits), Expect = 6.1e-13, Sum P(3) = 6.1e-13
 Identities = 37/119 (31%), Positives = 62/119 (52%)

Entry AB018340.1 from database TREMBL:
 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens mRNA for
 KIAA0797 protein, partial cds.
 Score = 540, P = 1.9e-50, identities = 120/243, positives = 155/243

Alert BLASTP hits for DKFZphfbr2_16g18, frame 3

TREMBL:ATT16L1_11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII
 project), N = 2, Score = 239, P = 2.1e-18

>TREMBL:ATT16L1_11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII project)
 Length = 710

HSPs:

Score = 239 (35.9 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 51/135 (37%), Positives = 78/135 (57%)

Query: 683 IVYPPPTKGGLGVTNEDLECEEGEFLNDVIIDFYLYLILEKASDELVERSHIFSSFF 742
 +VYP +V +D+E L+ F+ND IIDFY+KYL +S+ R H F+ FF
 Sbjct: 176 LVYPQGEPPDAVV-VRKQDIELLKPRRFINDTIIDFYIKYL-KNRISPKERGRFFHNCFF 233

Query: 743 YKCLTRKENNLTDNPNLSMAQRRHKRVRTWTRHINIFNKDYIFVPVNESSHWYLAVICF 802
 + RK NL + P+ + ++RV+ WT++++F KDYIF+P+N S HW L +IC
 Sbjct: 234 F----RKLANLDKGPSTCGGREAYQRVQKWKNVDLFEKDYIFIPINCSFHWSLVIICH 289

Query: 803 PWLEEAUYEDFFQTV 817
 P + + PQ V
 Sbjct: 290 PGELVPSHVENPQRV 304

Score = 70 (10.5 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 13/28 (46%), Positives = 15/28 (53%)

Query: 948 PIHLEKWFPRHVIKTKREDIRELILKLH 975
 P HL WFP KR +I EL+ LH
 Sbjct: 403 PSHLRNWFPKAEASLKRRNILELLYNLH 430

Pedant information for DKFZphfbr2_16g18, frame 3

Report for DKFZphfbr2_16g18.3


```

SEQ      MDKRLGRRPSSSEIITEGKRKKSSSLDSEIRKMLNAKPEDVHVHQSPLSKFRSSERWTLF
SEG      .....
PRD      cccccceccccceeeccccccccccchhhhhhhhhhhccccccccccccccccccccchh

SEQ      LQWERSLRNKVISLDHKNKKHIRGCPVTSRSSPERIPRVILTNLVGTTELGRKYIRTPPVT
SEG      .....
PRD      hhhhhhhhhhheeeccccceeeccccccccccccceeeeeeeeeccceeeccceeeccccc

SEQ      EGSLSDTDNLQSEQSLSSSSDGSLESYQNLNPHKSCYLSESGSQRSKTVDDNSAKQTAHNK
SEG      ..... xxxxxxxxxxxxxxxxxxxx .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhh

SEQ      EKRRKDDGISLLISDTQPEDLNSGRGCDHLEQESRNKDVKYSDSKVELTLISAKTKRRL
SEG      .....
PRD      hhhhhccccceeeccccccccccccccccccccccccccccccccccccceeeehhhhhhhh

SEQ      RNNLPDSQYCTSLDKSTEQTKKQEDDSTISTEFERPSENYHQDPKLPPEIITTKPKTSDF
SEG      .....
PRD      hccccccccccccccccchhhhhcccccccccccccccccccccccccccccccccccccc

SEQ      KLSSSLNSQELTSLNATKSASAGSTTETVEYSNSIDIVGISSLVEKDENELNTIEKPILRG
SEG      .....
PRD      cccccccccceehhhhhhhhhccccceeeccccceeeccccchhhhhhhhhhhcccccccc

SEQ      HNEGNSQLISAEPIVSSDDEEGPVEHKSSEILKLQSKQDRETTNENESTSESALLEPLI
SEG      ..... xxxxxxxxxxxxxxxxxxxx .....
PRD      cccccceeeccccceeeccccccccccchhhhhhhhhhhhhccccccccchhhhhcccce

SEQ      TCESVQMSSELCPYNPVMENISSIMPSNEMDLQDIFTSVYIGIKIGASKGCVTITKKY
SEG      .....
PRD      eccccccccccccccccccccceccccchhhhhheeeeeeeeeeeccccceeeeeeee

SEQ      IKIPFQVSLNEISLLVDTHLKRFLGWKSKDDNHSKRSHAILFFWVSSDYLQEIQTQLEH
SEG      .....
PRD      eeeccccceeeeeccccceeeeeeeccccccccceeeeeeecccchhhhhhhhhhh

SEQ      SVLSQQSKSSEFIFLELHNVPVSQREELKLKDIMTEISISGELELSYPLSWVQAFFLFQ
SEG      .....
PRD      hhhhccccceeeeeeeccccccchhhhhhhhhhhheeeccccceeeccccceeeeeeecc

SEQ      LSSKESSFIHYVCSTCSFPAGVAVAEEMKLKSVSQPSNTDAAKPTYTFLQKQSSGCGYSL
SEG      .....
PRD      cccccccccceeeccccccccchhhhhhhhhhhccccccccccccceeeccccccccce

SEQ      SITSNPDEEWREVHRHTGLVQKLIVYPPPTKGLGVTNEDLECLEEGEFLNDVIIDFYLK
SEG      .....
PRD      eccccccccceeeccccceeeeeeeccccccccccccccchhhhhhhhhccchhhhhhhhh

SEQ      YLILEKASDELVERSHIFSSFFYKCLTRKENNLTEDNPNLMAQRHRKRVRTWTRHINIF
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhh

SEQ      NKDYIFVPVNESSHWYLAVICFPWLEEAVYEDFPQTVSQQSQAQQSQSDNKTIDNDRLT
SEG      ..... xxxxxxxxxxxx .....
PRD      cceeeccccccccceeeccccchhhhhhhccccchhhhhhhhhcccccccccccccc

SEQ      STLSLSAEDSQSTESNMSVPKMKRKCILILDSLKAASVRNTVQNLRLEYLEVEWEVKLK
SEG      .....
PRD      cceeeccccccccceccccccccccccceeeccccccccchhhhhhhhhhhhhhhhhhh

SEQ      THRQFSKTNMVDLCPKVPKQDNSSDCGVYLLQYVESFFKDPIVNFELPIHLEKWFPRHVI

```



```

SEG .....
PRD hhhhhccccccccccccccccccccceeeehhhhhhhccccceccccccccccchhh

SEQ KTKREDIRELILKLHLQQQKSSS
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhcccccc

```

Prosites for DKF2phfbr2_16g18.3

PS00001	314->318	ASN_GLYCOSYLATION	PDOC00001
PS00001	365->369	ASN_GLYCOSYLATION	PDOC00001
PS00001	406->410	ASN_GLYCOSYLATION	PDOC00001
PS00001	440->444	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	600->604	ASN_GLYCOSYLATION	PDOC00001
PS00001	752->756	ASN_GLYCOSYLATION	PDOC00001
PS00001	759->763	ASN_GLYCOSYLATION	PDOC00001
PS00001	790->794	ASN_GLYCOSYLATION	PDOC00001
PS00001	830->834	ASN_GLYCOSYLATION	PDOC00001
PS00001	856->860	ASN_GLYCOSYLATION	PDOC00001
PS00001	922->926	ASN_GLYCOSYLATION	PDOC00001
PS00004	8->12	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	21->25	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	162->165	PKC_PHOSPHO_SITE	PDOC00005
PS00005	172->175	PKC_PHOSPHO_SITE	PDOC00005
PS00005	233->236	PKC_PHOSPHO_SITE	PDOC00005
PS00005	236->239	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	291->294	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	515->518	PKC_PHOSPHO_SITE	PDOC00005
PS00005	562->565	PKC_PHOSPHO_SITE	PDOC00005
PS00005	602->605	PKC_PHOSPHO_SITE	PDOC00005
PS00005	747->750	PKC_PHOSPHO_SITE	PDOC00005
PS00005	874->877	PKC_PHOSPHO_SITE	PDOC00005
PS00005	879->882	PKC_PHOSPHO_SITE	PDOC00005
PS00005	901->904	PKC_PHOSPHO_SITE	PDOC00005
PS00005	962->965	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	167->171	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	225->229	CK2_PHOSPHO_SITE	PDOC00006
PS00006	251->255	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	396->400	CK2_PHOSPHO_SITE	PDOC00006
PS00006	402->406	CK2_PHOSPHO_SITE	PDOC00006
PS00006	408->412	CK2_PHOSPHO_SITE	PDOC00006
PS00006	488->492	CK2_PHOSPHO_SITE	PDOC00006
PS00006	509->513	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	562->566	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	638->642	CK2_PHOSPHO_SITE	PDOC00006
PS00006	664->668	CK2_PHOSPHO_SITE	PDOC00006
PS00006	697->701	CK2_PHOSPHO_SITE	PDOC00006
PS00006	747->751	CK2_PHOSPHO_SITE	PDOC00006
PS00006	826->830	CK2_PHOSPHO_SITE	PDOC00006
PS00006	846->850	CK2_PHOSPHO_SITE	PDOC00006
PS00006	962->966	CK2_PHOSPHO_SITE	PDOC00006
PS00007	216->223	TYR_PHOSPHO_SITE	PDOC00007
PS00008	84->90	MYRISTYL	PDOC00008
PS00008	106->112	MYRISTYL	PDOC00008
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	468->474	MYRISTYL	PDOC00008

PS00008	505->511	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00008	693->699	MYRISTYL	PDOC00008
PS00009	6->10	AMIDATION	PDOC00009
PS00009	18->22	AMIDATION	PDOC00009
PS00009	109->113	AMIDATION	PDOC00009

(No Pfam data available for DKF2phfbr2_16g18.3)

DKFZphfbr2_16112

group: transmembrane protein

DKFZphfbr2_16112 encodes a novel 185 amino acid protein, with strong similarity to PUT2 protein of Fugu rubripes.

The novel protein contains 1 transmembrane region.

PUT 2 is a Fugu rupies protein similar to the neural cell adhesion molecule L1 (L1-CAM) a mitosis-specific chromosome segregation protein (SMC1) and the calcium channel alpha-1 subunit homolog (CCA1).

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

strong similarity to Fugu rubripes PUT2

complete cDNA, complete cds, EST hits,
TRANSMEMBRANE 1

Sequenced by LMU

Locus: /map="873.3/875.1 cR from top of Chr1 linkage group"

Insert length: 1552 bp

Poly A stretch at pos. 1528, polyadenylation signal at pos. 1506

```

1  GGGGGGGGAC AACTGGGTCT TTGCGGCTG CAGCGGGCTT GTAGGCGTCC
51 GGCTTTGCTG GCCCAGCAAG CCTGATAAGC ATGAAGCTCT TATCTTTGGT
101 GGCTGTGGTC GGGTGTTCG TGGTGCCCC AGCTGAAGCC AACAAAGATT
151 CTGAAGATAT CCGGTGCAAA TGCATCTGTC CACCTTATAG AACATCAGT
201 GGGCACATTT ACAACCAGAA TGTATCCCAG AAGGACTGTT GTAGCAACTG
251 CCTGCACGTG GTGGAGCCCA TGCCAGTGCC TGGCCATGAC GTGGAGGCCT
301 ACTGCCTGCT GTGCGAGTGC AGGTACGAGG AGCGCAGCAC CACCACCATC
351 AAGGTCATCA TTGTCACTA CCTGTCCGTG GTGGGTGCCG TGTGTCTCTA
401 CATGGCCTTC CTGATGCTGG TGGACCCTCT GATCCGAAAG CCGGATGCAT
451 ACACGTAGCA ACTGCACAAT GAGGAGGAGA ATGAGGATGC TCGCTCTATG
501 GCAGCAGCTG CTGCATCCCT CGGGGGACCC CGAGCAAACA CAGTCTTGGA
551 GCGTGTGGAA GGTGCCCAGC AGCGGTGGAA GCTGCAGGTG CAGGAGCAGC
601 GGAAGACAGT CTTGATCGG CACAAGATGC TCAGCTAGAT GGGCTGGTGT
651 GGTGGGTGTA AGGCCCCAAC ACCATGGCTG CCAGCTTCCA GGCTGGACAA
701 AGCAGGGGGC TACTTCTCCC TTCCCTCGGT TCCAGTCTTC CCTTTAAAG
751 CCTGTGGCAT TTTTCTCCT TCTCCCTAAC TTAGAAATG TTGTACTTGG
801 CTATTTTGAT TAGGGAAGAG GGATGTGGTC TCTGATCTCT GTTGTCTTCT
851 TGGGTCTTTG GGGTTGAAGG GAGGGGGAAG GCAGGCCAGA AGGGAATGGA
901 GACATTTCAG GCGGCCCTAG GAGTGGATGC GATCTGTCTC TCCTGGCTCC
951 ACTCTTGCCG CCTTCCAGCT CTGAGTCTTG GGAATGTTGT TACCCTTGGA
1001 AGATAAAGCT GGGTCTTCAG GAACTCAGTG TTTGGGAGGA AAGCATGGCC
1051 CAGCATTTCAG CATGTGTTCC TTCTGCAGT GGTCTTATC ACCACCTCCC
1101 TCCAGCCCC AGCGCCTCAG CCCCAGCCCC AGCTCCAGCC CTGAGGACAG
1151 CTCTGATGGG AGAGCTGGGC CCCCTGAGCC CACTGGGTCT TCAGGGTGCA
1201 CTGGAAGCTG GTGTTGCTG TCCCCTGTGC ACTTCTCGCA CTGGGGCATG
1251 GAGTGCCCAT GCATACTCTG CTGCCGGTCC CCTCACCTGC ACTTGAGGGG
1301 TCTGGGCGAG CCCTCCTCTC CCCAGTGTCC ACAGTCACTG AGCCAGACGG
1351 TCGGTTGGAA CATGAGACTC GAGGCTGAGC GTGGATCTGA ACACCACAGC
1401 CCCTGTACTT GGGTTGCCTC TTGTCCCTGA ACTTCGTGT ACCAGTGCAT
1451 GGAGAGAAAA TTTTGTCCCT TTGTCTTAGA GTTGTGTGTA AATCAAGGAA
1501 GCCATCATTA AATTGTTTTA TTTCTCTCAA AAAAAAAAAA AAAAAAATA
1551 TC

```

BLAST Results

Entry HS808349 from database EMBL:

human STS WI-11986.

Score = 1716, P = 5.7e-73, identities = 364/378

Entry HS487355 from database EMBL:

human STS WI-13088.

Score = 1358, P = 1.3e-56, identities = 274/277

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 81 bp to 635 bp; peptide length: 185
 Category: similarity to unknown protein

1 MKLLSLVAVV GCLLVPPAEA NKSSDIRCK CICPPYRNIS GHIYNQNVSO
 51 KDCCSNCLHV VEPMPVPGHD VEAYCLLCEC RYEERSTTTI KVIIIVYLSV
 101 VGALLLYMAF LMLVDPLIRK PDAYTEQLHN EEENEDARSM AAAASLGGP
 151 RANTVLERVE GAQQRWKLQV QEQRKTVFDR HKMLS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16i12, frame 3

TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
 rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene,
 complete cds; putative protein 1 (PUT1) gene, partial cds;
 mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
 gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1)
 and putative protein 2 (PUT2) genes, partial cds, complete sequence., N
 = 1, Score = 655, P = 2.8e-64

TREMBL:CER12C12_5 gene: "R12C12.6"; Caenorhabditis elegans cosmid
 R12C12., N = 1, Score = 225, P = 1e-18

>TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
 rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete
 cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific
 chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and
 calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
 (PUT2) genes, partial cds, complete sequence.
 Length = 187

HSPs:

Score = 655 (98.3 bits), Expect = 2.8e-64, P = 2.8e-64
 Identities = 124/163 (76%), Positives = 140/163 (85%)

Query: 22 KSSDIRCKCICPPYRNISGHIYNQNVSKDCCSNCLHVVEPMPVPGHDVEAYCLLCECR 81
 KS +D+RCKCICPPYRNISGHIYN+N +QKDC NCLHVV+PMPVPG+DVEAYCLLCEC+
 Sbjct: 31 KSFDDVRCKCICPPYRNISGHIYNRNFTQKDC--NCLHVVDPMPVPGNDVEAYCLLCECK 88

Query: 82 YEERSTTTIKVIIIVYLSVVGALLLYMAFLMLVDPLIRKPDAYTEQLHNEENEDARSM 141
 YEERST TI+V I+I+LSVVGALLLYM FL+LVDPLIRKPD + LHNEE++ED +
 Sbjct: 89 YEERSTNTIRVTIIIFLSVVGALLLYMLFLLLVDPPLIRKPDPLAQLHNEEDSEDIQPQM 148

Query: 142 AAAASLGGP-RANTVLERVEGAQQRWKLQVQEQRKTVFDRHKML 184
 + G P R NTVLERVEGAQQRWK QVQEQRKTVFDRHKML
 Sbjct: 149 S-----GDPARGNTVLERVEGAQQRWKKQVQEQRKTVFDRHKML 187

Pedant information for DKFZphfbr2_16i12, frame 3

Report for DKFZphfbr2_16i12.3

[LENGTH] 185
 [MW] 20764.29
 [PI] 6.21
 [HOMOL] TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes
 neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1
 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
 gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
 (PUT2) genes, partial cds, complete sequence. 3e-68
 [PROSITE] MYRISTYL 1
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 2
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] SIGNAL_PEPTIDE 21


```

[KW]          TRANSMEMBRANE 1
[KW]          LOW_COMPLEXITY 2.70 %

SEQ  MKLLSLVAVVGCLLVPPAEANKSSEDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHV
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  VEPMPVPGHDVEAYCLCECRYEERSTTTIKVIIIVYLSVVGALLLYMAFLMLVDPLIRK
SEG  .....
PRD  eccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  PDAYTEQLHNEEENEDARSMAAAAASLGGPRANTVLERVEGAQQRWKLQVQEQRKTVFDR
SEG  .....xxxxx.....
PRD  cchhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhchhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  HKMLS
SEG  .....
PRD  hhccc
MEM  .....

```

Prosites for DKFZphfbr2_16i12.3

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	38->42	ASN_GLYCOSYLATION	PDOC00001
PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	148->154	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16i12.3)

DKFZphfbr2_16k22

group: brain derived

DKFZphfbr2_16k22 encodes a novel 108 amino acid protein with very weak similarity to thioredoxin of *Bacillus subtilis*.

No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to thioredoxin

complete cDNA, complete cds, genomic DNA?
no EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2088 bp

Poly A stretch at pos. 2065, no polyadenylation signal found

```
1 AAAAGGAAGA AGGAAATAAG GATATTTCAA GGGTTACCAA AGTCGAGGAA
51 AACTATTTTA AGAAGAAATC TGAATTATTT GTGCACATAG GTTGTAATAA
101 TAGCATCTTG CATTAAATGG TGTTTTCTAG CTTACAAAGT GGATTCATAT
151 ACACATATTGT AACTGACTCT CTACAACTT GCAAGGTTAG CAAGACAAAT
201 GGTATTTTAA GATAACAAAC TGAGACTCAA AAAAGGCAAG TAACTCGTTC
251 TACTTCCCAA AGCCAGAAAG TGGCAAAATA GAAAATGGAT CCTGAATCTC
301 CAACACCATG CAAACTAAGA GAGGGAATCC TCTGTAGAGG GAATGGAAGT
351 AAAAAGGCAC AAGTGGTGAT GTCACCTTCT GAACAGAGAT GGAACTTTTT
401 TTCCTCTGAG AAAAAAGAGA AAAGATAGTT TTAAGTGGCA AAAGAACATG
451 AAGCAATGTG AGGTGAAGAA ACAGAAAGA CTATGGATGG AATTCCTAGA
501 TGTGAGATAC ACAAAGTTCC ATTTCAAAGA GAAATATCTA TAGATAGGCA
551 TAAAGTTTACA CACCTGAACT ACCAACTCTG AACCAGTAAC TCAAGAGATA
601 TTTTGTGTGT CCCACAAGCC ATATGGCTCT GGGGACAAAT TATCTGAAAG
651 TGCCCAATAA GAAAAATATT TGAGGAAGGG GAGTTGGTGA GTGAATGAAT
701 TAAAGGACAT CAGAAAGATA CATTGACTGT TCTCCTTCCC AGGAAACAAA
751 GTGGCTAAGT CAAACAACG GGCAGCTGTG GGATAGCAA GAAAAAATAA
801 CTTCAGGCC CAGGTTCTAG TGAAGCTAC TATGGAAGTT AGCCACTCAA
851 CTTTAGAACC AGAGGCTTCT TTTCTCCTC CTTCTTATC TTTTCTAGT
901 TATAGCAAAT TTATATTGAG CCACTTATTC TTTCTGAATG CTAGTTCCCC
951 TTTAGCATTT CTTTTTCTT ATTCCCTTG GACTGGCCCA ATGCTTTGGC
1001 CCCTTATCAA AGCATTCTT AAGAAACAGT CTGACAGCTC TAATTTGCAT
1051 CTGGTTATGC AAGATGTGGT TAAGAACATG GACTCTGGAG GTAAATACAC
1101 CTTGATTCCA ATTCATTCTC TCATTTATTC ATTCAGCAA TATTTAGTGA
1151 ACATCTAACA TGTGCTAGGC ACTGTTCTAG TTGCTGAGGA TACAGCTTCA
1201 AACAAAATAA GGTCTCTGCA AGGATGCCTT CTCTTACCAC TCCTATTGAG
1251 CGTAGTATTG GAAGTCCTGG CCAGGGCAAT CAGGCAAGAA AAAGAAATCA
1301 AGGTCATCCA AATAGGAAGA GAGGAAGTCA AACTATCCCT GTTTACAGAC
1351 AACATGATCC TACATCTAGA AAAAAACCA TTGTCTTAGC CAAAAGCTT
1401 CTTAGGCTGA TAAACAACTT CAGCAAAGTC TTAGGATACA AAATCCATGT
1451 GCAAAAAACA CTAGCATCTT TATACACCAA CAACAGTCAA GCCGAGATCC
1501 AAATCAGGAA CAAACTCCTA TTCACAATTG CCACAAAAC AATAGAACAG
1551 GAAAAACAGCT AACTAGGAAG GTGAAAGATC TCTACAAGGA GAACTACAAA
1601 CCACTGCTCA CAGAAATCAG AGATGACACA TATAAATGGA AAAACATTCC
1651 ATGATCATGG ATAGGAAGAA TGAATATTAC TGAATGGCT ATACTGTCCA
1701 AAGCAATTTA TAGATTCAAT GCTATTCTTA GTAACTACC ATTGAGATTT
1751 TTTACAGAAC TAGAAAAAAA AAAAATCTT TTAAGGCTGG GCGCAGTGGC
1801 TCTCACCTGT AATCCCAGCA CTTTGGGAGG CCGAGATGGG TGGATCACGA
1851 GGTGAGGAGA TGGAAACAT CCTGGCTAAC ATGGTGAAAC CCCGCTCTTA
1901 CTAAAAATAC AAAAAATTAG CCAGGCGTGG TGGTGGGCGC CTGTAATCCC
1951 AGCTGCTCGG GAGGCTGAGG CAGGATAATG GTGTGAACCC GGGAGGCAGA
2001 GCTTGCAAGT AGCTGAGATT GCACCACTGC ACTCCAGCCT GAGGGACAGA
2051 GTGAGACTCC ATCTCAAAAA AAAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 832 bp to 1155 bp; peptide length: 108
Category: putative protein

1 MEVSHSTLEP EASFPPPFSL FLVYSKFILS HLFFLNASSP LAFLFLHSLW
51 TGPMLWPLIK AFSKKQSDSS NLHLVMQDVV KNMDSGGKYT LIPIHSLIYS
101 FSKYLVNI

BLASTP hits

Entry B37192 from database PIR:
thioredoxin - Bacillus subtilis Score = 71 (25.0 bits), Expect = 0.040,
P = 0.039
Identities = 16/49 (32%), Positives = 30/49 (61%)

Alert BLASTP hits for DKFZphfbr2_16k22, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_16k22, frame 1

Report for DKFZphfbr2_16k22.1

[LENGTH] 108
[MW] 12281.47
[pI] 8.06
[PROSITE] MYRISTYL 1
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MEVSHSTLEPEASFPPPFSLFLVYSKFILSHLFFLNASSPLAFLFLHSLWTGPMLWPLIK
PRD cccccccccccccccccchhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhh
SEQ AFSKKQSDSSNLHLVMQDVVKNMDSGGKYTLIPIHSLIYSFSKYLVINI
PRD hhhccccccccceehhhhhhhccccccccceeeccceeecccccccccc

Prosites for DKFZphfbr2_16k22.1

PS00001	36->40	ASN_GLYCOSYLATION	PDOC00001
PS00004	64->68	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00006	6->10	CK2_PHOSPHO_SITE	PDOC00006
PS00008	86->92	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16k22.1)

DKFZphfbr2_16112

group: transmembrane protein

DKFZphfbr2_16112 encodes a novel 267 amino acid protein with similarity to gallus gallus putative transmembrane protein E3-16

The novel protein contains one putative transmembrane domain. In chicken, E3-16 is expressed specifically in the inner ear.

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neurons involved in perception of hearing.

similarity to gallus putative transmembrane protein E3-16

complete cdna, complete cds, EST hits
potential start at Bp 73 matches kozak consensus PyCCatag
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 2042 bp

Poly A stretch at pos. 2024, polyadenylation signal at pos. 2003

```
1 GGGGGCGGCG GAGGCAGAGA CCGAGGCTGC ACCGGCAGAG GCTGCGGGGC
51 GGACGCGCGG GCCGGCGCAG CCATGGTGAA GATTAGCTTC CAGCCGCGCG
101 TGGCTGGCAT CAAGGGCGAC AAGGCTGACA AGGCGTCGGC GTCGGCCCCCT
151 GCGCGCGCCT CGGCCACCGA GATCCTGCTG ACGCCGGCTA GGGAGGAGCA
201 GCGCCCAACA CATCGATCCA AGAGGGGGGG CTCAGTGGGC GCGGTGTGCT
251 ACCTGTGCGT GGGCATGGTC GTGCTGCTCA TGGGCTCGT GTTCGCCTCT
301 GTCTACATCT ACAGATACTT CTTCCTTGCG CAGCTGGCCC GAGATAACTT
351 CTTCGGCTGT GGTGTGCTGT ATGAGGACTC CCTGTCTCC CAGGTCCGGA
401 CTCAGATGGA GCTGGAAGAG GATGTGAAAA TCTACCTCGA CGAGAACTAC
451 GAGCGCATCA ACGTGCCTGT GCGCCAGTTT GCGCGCGGTG ACCCTGCAGA
501 CATCATCCAT GACTTCCAGC GGGGTCTGAC TGCCTACCAT GATATCTCCC
551 TGGACAAGTG CTATGTCATC GAACTCAACA CCACCATTTG GCTGCCCCCT
601 CGCAACTTCT GGGAGCTCCT CATGAACGTG AAGAGGGGGA CCTACCTGCC
651 GCAGACGTAC ATCATCCAGG AGGAGATGGT GGTACGGGAG CATGTCACTG
701 ACAAGGAGGC CCTGGGGTCC TPCATCTACC ACCTGTGCAA CGGGAAGAC
751 ACCTACCGGC TCCGGCGCCG GCAACGCGG AGGCGGATCA ACAAGCGTGG
801 GGCCAAGAAG TGCAATGCCA TCCGCCACTT CGAGAACACC TTCGTGGTGG
851 AGACGCTCAT CTGCGGGGTG GTGTGAGGCC CTCCTCCCCC AGAACCCCCCT
901 GCGGTGTTCC TCTTTTCTTC TTTCCGGCTG CTCTCTGGCC CTCTCTCTTC
951 CCCCTGCTTA GCTTGTACTT TGGACGCGTT TCTATAGAGG TGACATGTCT
1001 CTCCATTCTT CTCCAACCTT GCCCACCTCC CTGTACCAGA GCTGTGATCT
1051 CTCGGTGGGG GGCCCATCTC TGCTGACCTG GGTGTGGCGG AGGGAGAGGC
1101 GATGCTGCAA AGTGTCTTCT GTGTCCCACT GTCTTGAAGC TGGGCTTGCC
1151 AAAGCCTGGG CCCACAGCTG CACCGGCAGC CCAAGGGGAA GGACCGGTTG
1201 GGGGAGCCGG GCATGTGAGG CCCTGGGCAA GGGGATGGGG CTGTGGGGGG
1251 GGGGCGGCAT GGGCTTCAGA AGTATCTGCA CAATTAGAAA AGTCTCTAGA
1301 AGCTTTTCTT TGGAGGGTAC ACTTTCTTCA CTGTCCCTAT TCCTAGACCT
1351 GGGGCTTGAG CTGAGGATGG GACGATGTGC CCAGGGAGGG ACCCACCAGA
1401 GCACAAGAGA AGGTGGCTAC CTGGGGGTGT CCCAGGACT CTGTCACTGC
1451 CTTACGCCCA CCAGCAGGAG CTGGAGTTT GGGGAGTGGG GATGAGTCCG
1501 TCAAGCACAA CTGTTCTCTG AGTGGAACCA AAGAAGCAAG GAGCTAGGAC
1551 CCCCAGTCCT GCGCCCCAGG AGCACAAGCA GGGTCCCCCT AGTCAAGGCA
1601 GTGGGATGGG CGGCTGAGGA ACGGGGCAGG CAAGGTCACT GCTCAGTCAC
1651 GTCCACGGGG GACGAGCCGT GGGTCTGCT GAGTAGGTGG AGCTCATTGC
1701 TTTCTCCAAG CTTGGAACGT TTTTGAAAGA TAACACAGAG GGAAGGGGAG
1751 AGCCACCTGG TACTTGTTCA CCCTGCCTCC TCTGTTCTGA AATTCCATCC
1801 CCCTCAGCTT AGGGGAATGC ACCTTTTCTC CTTTCTTCT CACTTTTGCA
1851 TGTCTTTTCT GATCATTCTG TATGCTAACC GTTCTCAGCC CTGAGCCTTG
1901 GAGAGGAGGG CTGTAACGCC TTCAGTCAGT CTCTGGGGAT GAAACTCTTA
1951 AATGCTTTGT ATATTTTCTC AATTAGATCT CTTTTCAGAA GTGTCTATAG
2001 AACAATAAAA ATCTTTTACT TCTGAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation
using cDNA library subtraction. Molecular cloning and
characterization of a gene belonging to a novel multigene
family of integral membrane proteins.

Peptide information for frame 1

ORF from 73 bp to 873 bp; peptide length: 267
Category: similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGGSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VVPVQFGGGD PADIIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRFNW ELLMNVKRGY YLPQTYIIQE
201 EMVVTETVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16112, frame 1

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16)., N = 1, Score = 573, P = 1.4e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N =
1, Score = 559, P = 4.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1,
Score = 452, P = 9.1e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16). Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.4e-55, P = 1.4e-55
Identities = 118/264 (44%), Positives = 175/264 (66%)

```

Query: 1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY 60
      MVK+SF A+A + A+K ++ ++L+ P + + P+ G C+
Sbjct: 1 MVKVSFNSALA--HKEAANKKEENS-----QVLILPP-DAKEPEDVVVPAGHKRAWCW 50

Query: 61 -LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL-----SQVRTQM- 112
      + G+ +L G++ Y+Y+YF Q + CG+ Y ED LS +Q+++
Sbjct: 51 CMCFLAFLMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKIEDGLSLPESGAQLKSARY 107

Query: 113 -ELEEDVKIYLDENYERINVVPVQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTT 171
      +E++++I +E+ E I+VPVP+F DPADI+HDF R LTAY D+SLDKCYVI LNT+
Sbjct: 108 HTIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRLTAYLDLSLDKCYVIPLNTS 167

Query: 172 IVLPPRFNWELLNMVNRGTYPQTYIIQEEMVVTETVSDKEALGSFIYHLCNGKDTYRLR 231
      +V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+
Sbjct: 168 VVMPPKNFLELLINIKAGTYLPQSYLIHEQMIIVTDRIENVVDQLGFFIYRLCRGKETYKLQ 227

Query: 232 RRATRRRINKRGAKNCNAIRHFENTFVVETLIC 264
      R+ + I KR A NC IRHFEN F +ETLIC
Sbjct: 228 RKEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

```

Pedant information for DKFZphfbr2_16112, frame 1

Report for DKFZphfbr2_16112.1

[LENGTH] 267
[MW] 30223.94


```

SEQ      MVKISFQPAVAGIKGDKADKASAPASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      cccccccccchhhhhhhhhhhhhhhhhcccccceccccccccccccccccccccchh
MEM      .....MMMMMMMMMM.....

SEQ      LSMGMVLLMLGLVFASVYIYRYFFLAQLARDNFFRCGLVYEDSLSSQVRTQMELEEDVKI
SEG      ..XXXXXXXXXXXXX.....
PRD      hhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhccceeeeeeccccccccchhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMMMMMMM.....

SEQ      YLDENYERINVPVPQFGGGDPADIHDFQRLTAYHDISLDKCYVIELNTTIVLPPRNF
SEG      .....
PRD      hccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhccceeeeeeccccchh
MEM      .....

SEQ      ELLMNVKRGTYLPQTYIIQEEMVUTEHVSDEALGFSFIYHLCNGKDTYRLRRATRRRIN
SEG      .....XXXXXXXXXXXXX.....
PRD      hhhhhcccccceeeehhhhhhhccccchhhhhheeeccccchhhhhhhhhhhhhhh
MEM      .....

SEQ      KRGAKNCNAIRHFENTFVVETLICGVV
SEG      xx.....
PRD      hhhhhccceeeccccchhhhhheeeccc
MEM      .....

```

PS000001	169->173	ASN_GLYCOSYLATION	PDOC000001
PS000004	187->191	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	232->236	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	49->52	PKC_PHOSPHO_SITE	PDOC000005
PS000005	209->212	PKC_PHOSPHO_SITE	PDOC000005
PS000005	227->230	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000006	30->34	CK2_PHOSPHO_SITE	PDOC000006
PS000006	110->114	CK2_PHOSPHO_SITE	PDOC000006
PS000006	209->213	CK2_PHOSPHO_SITE	PDOC000006
PS000007	119->127	TYR_PHOSPHO_SITE	PDOC000007
PS000008	52->58	MYRISTYL	PDOC000008
PS000008	53->59	MYRISTYL	PDOC000008
PS000008	71->77	MYRISTYL	PDOC000008
PS000008	138->144	MYRISTYL	PDOC000008
PS000008	243->249	MYRISTYL	PDOC000008
PS00294	264->268	PRENYLATION	PDOC00266

151

DKFZphfbr2_22f21

group: brain derived

DKFZphfbr2_22f21 encodes a novel 567 amino acid protein with weak similarity to C. elegans cosmid C18C4.5

No informative BLAST results; no predictive prosite, pfam or SCOP motifs

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to C.elegans C18C4.5

EST HSAA6531/HSAA5273/ defines splice variant, or unspliced cDNA additional -180 Bp at position 250

Sequenced by AGOWA

Locus: /map="311.4 cR from top of Chr14 linkage group"

Insert length: 1910 bp

Poly A stretch at pos. 1887, polyadenylation signal at pos. 1867

```
1 TGGGCCCTTA GCAACGGCCT GGCACGGTT TCCTTGCTGC TGCAGCCCCC
51 GTCGGGTCCT CTTTCCAGT CCTCCACTGC CGGGGCTGGG CCCGGCCGCG
101 GGAAGGACCG AAGGGGATAC AGCGTGTCCC TCGGGCGGCT GCAAGAGGAC
151 TAAGCATGGA TGGCAGCCGG AGAGTCAGAG CAACCTCTGT CCTTCCCAGA
201 TATGGTCCAC CGTGCCTATT TAAAGGACAC TTGAGCACCA AAAGTAATGC
251 TGCAGTAGAC TGCTCGGTTT CAGTAAGCAT GAGTACCAGC ATAAAGTATG
301 CAGACCAACA ACGAAGAGAG AAACCTCAAA AGGAATTAGC ACAATGTGAA
351 AAAGAGTTCA AATTAATAA AACTGCAATG CGAGCCATT ATAAAAATAA
401 TTCCAAGTCA CTTTTAATA CCTTACAAGA GCCCTCAGGC GAACCGCAAA
451 TTGAGGATGA CATGTTAAAA GAAGAAATGA ATGGATTTC ATCCTTTGCA
501 AGGTCACTAG TACCTCTTTC AGAGAGACTA CACCTAAGTC TACATAAATC
551 CAGTAAAGTC ATCACAATG GTCCTGAGAA GAACTCCAGT TCCTCCCCGT
601 CCAGTGTGGA TTATGCAGCC TCCGGGCCCC GAAACTGAG CTCTGGAGCC
651 CTGTATGGCA GAAGGCCAGC AAGCACATTC CCAAAATCCC ACCGGTTTCA
701 GTTAGTCATT TCGAAAGCAC CCAGTGGGGA TCCTTTGGAT AAACATTCTG
751 AACTCTTTTC TAACAAACAA TTGCCATTCA CTCCTCGCAC TTTAAAAACA
801 GAAGCAAAAT CTTCTCTGTC ACAGTATCGC TATTATACAC CTGCCAAAAG
851 AAAAAAGGAT TTTACAGATC AACGGATAGA AGCTGAAACC CAGACTGAAT
901 TAAGCTTTAA ATCTGAGTTG GGGACAGCTG AGACTAAAAA CATGACAGAT
951 TCAGAAATGA ACATAAAGCA GGCATCTAAT TGTGTGACAT ATGATGCCAA
1001 AGAAAAAATA GCTCCTTTAC CTTTAGAAGG GCATGACTCA ACATGGGATC
1051 AGATTAAAGG TGATGCTCTT CAGCATTCCT CACCAAGGGC AATGTGTCAG
1101 TATTCCTGTA AGCCCCCTTC AACTCGTAAA ATCTACTCTG ATGAAGAAGA
1151 ACTGTTGTAT CTGAGTTTCA TTGAAGATGT AACAGATGAA ATTTTGAAAC
1201 TTGGTTTATT TTCAAACAGG TTTTGTAGAA GACTGTTTGA GCGACATATA
1251 AAACAAAATA AACATTTGGA GGGGGAAAAA ATGCGCCACC TGCTGCATGT
1301 CCTGAAAGTA GACTTAGGCT GCACATCGGA GGAAGACTCG GTAAAGCAAA
1351 ATGATGTTGA TATGTTGAAT GTATTGATT TTGAAAAGGC TGGGAATTCA
1401 GAACCAATAA AATTAATAAA TGAAAGTGAA GTAACAATTC AGCAGGAACG
1451 TCAACAATAC CAAAAGGCTT TGGATATGTT ATTGTCGGCA CCAAAGGATG
1501 AGAACGAGAT ATCCCTTCA CCAACTGAAT TTTTCATGCC TATTTATAAA
1551 TCAAAGCATT CAGAAGGGGT TATAATTCAA CAGGTGAATG ATGAAACAAA
1601 TCTTGAACCT TCAACTTTGG ATGAAATCA TCCAAGTATT TCAGACAGTT
1651 TAACAGATCG GGAACCTTCT GTGAATGTCA TTGAAGGTGA TAGTGACCTT
1701 GAAAAGGTTG AGATTTCAAA TGGATTATGT GGTCTTAACA CATCACCTTC
1751 CCAATCTGTT CAGTTCTCCA GTGTCAAAGG CGACAATAAT CATGACATGG
1801 AGTTATCAAC TCTTAAATC ATGGAATGA GCATTGAGGA CTGCCCTTTG
1851 GATGTTTAAT CTTCAATTAAT AAATACCTCA AATGGCCAGT AAAAAAATAA
1901 AAAAAAATAA
```

BLAST Results

Entry HS477360 from database EMBL:

human STS WI-14643.

Length = 418

Minus Strand HSPs:

Score = 1850 (277.6 bits), Expect = 2.5e-77, P = 2.5e-77

Identities = 392/405 (96%), Positives = 392/405 (96%), Strand = Minus /

Plus

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 1856 bp; peptide length: 567
 Category: similarity to unknown protein

```

1 MDGSRVRAT SVLPYRGPPC LFKGHLSTKS NAAVDCSVPV SMSTSIKYAD
51 QQRREKLKKE LAQCEKEFKL TKTAMRANYK NNSKSLFNTL QEPSGEPQIE
101 DDMLKEEMNG FSSFARSLVP SSERLHLSLH KSSKVITNGP EKNSSSSPSS
151 VOYAASGPRK LSSGALYGRR PRSTFPNSHR FQLVISKAPS GDLLDKHSEL
201 FSNKQLPFTT RTLKTEAKSF LSQYRYTTPA KRKKDFTDQR IEAETQTELS
251 FKSELGTAET KNMTDSEMNI KQASNCVTYD AKEKIAPLPL EGHDSWDEI
301 KDDALQHSSP RAMCQYSLKP PSTRKIYSDE EELLYLSFIE DVTDEILKLG
351 LFSNRFLERL FERHIKQNKH LECEKMRHLL HVLKVDLGCT SEENSVKQND
401 VDMLNVDFE KAGNSEPNKL KNESEVTIQQ ERQQYQKALD MLLSAPKDEN
451 EIFPSPTEFF MPIYKSKHSE GVIIQQVNDE TNLETSTLDE NHPSISDSLT
501 DRETSVNVIE GDSDPKVEI SNGLCGLNTS PSQSVQFSSV KGDNNHDMEL
551 STLKIMEMSI EDCPLDV

```

BLASTP hits

Entry CEC18C4_3 from database TREMBL:
 "C18C4.5"; Caenorhabditis elegans cosmid C18C4.
 Length = 1091
 Score = 98 (34.5 bits), Expect = 0.29, P = 0.25
 Identities = 105/470 (22%), Positives = 192/470 (40%)

Alert BLASTP hits for DKFZphfbr2_22f21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22f21, frame 3

Report for DKFZphfbr2_22f21.3

```

[LENGTH]      567
[MW]           64120.02
[pI]           5.68
[PROSITE]      AMIDATION      1
[PROSITE]      MYRISTYL       3
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 16
[PROSITE]      PKC_PHOSPHO_SITE 18
[PROSITE]      ASN_GLYCOSYLATION 4
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 1.23 %

```

```

SEQ  MDGSRVRATSVLPYRGPPCLFKGHLSTKSNAAVDCSVPVSMSTSIKYADQQRREKLKKE
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LAQCEKEFKLTKTAMRANYKNNSKSLFNTLQEPSGEPQIEDDMLKEEMNGFSSFARSLVP
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc

SEQ  SSERLHLSLHKSSKVITNGPEKNSSSSPSSVDYAASGPRKLSSGALYGRRPRSTFPNSHR
SEG  .....xxxxxxx.....
PRD  ccchhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  FQLVISKAPSGDLLDKHSELFSNKQLPFTRTLKTEAKSFLSQYRYTTPAKRKKDFTDQR
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  IEAETQTELSFKSELGTAETKNMTDSEMNIKQASNCVTYDAKEKIAPLPLEGHDSWDEI
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc

```



```

SEQ      KDDALQHSSPRAMCQYSLKPPSTRKIYSDEEELLYLSFIEDVTDEILKGLFSNRFLERL
SEG      .....
PRD      cccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhccchhhhhhh

SEQ      FERHIKQNKHLEGEKMRHLLHVLKVDLGCTSEENSVKQNDVMDLNVDFEKAGNSEPNKL
SEG      .....
PRD      hhhhhhhhhhhccchhhhhhhhhccccccccccccccccccccceeecccccccccccc

SEQ      KNESEVTIQERQQYQKALDMLLSAPKDENEIFPSPTEFFMPIYKSKHSEGVIIQQVNDE
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeecccc

SEQ      TNLETSTLDENHPSISDSLTDRETSVNVIEGSDPEKVEISNGLCGLNTSPSQSVQFSSV
SEG      .....
PRD      cccccccccccccccccccccccccccccceeeccccccccccccccccccccccccceeecc

SEQ      KGDNNHDMELSTLKIMEMSIEDCLDV
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhcccccc

```

Prosites for DKFZphfbr2_22f21.3

PS00001	81->85	ASN_GLYCOSYLATION	PDOC00001
PS00001	143->147	ASN_GLYCOSYLATION	PDOC00001
PS00001	262->266	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00004	159->163	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00005	178->181	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	212->215	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	317->320	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	353->356	PKC_PHOSPHO_SITE	PDOC00005
PS00005	395->398	PKC_PHOSPHO_SITE	PDOC00005
PS00005	500->503	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	264->268	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	481->485	CK2_PHOSPHO_SITE	PDOC00006
PS00006	486->490	CK2_PHOSPHO_SITE	PDOC00006
PS00006	494->498	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	500->504	CK2_PHOSPHO_SITE	PDOC00006
PS00006	513->517	CK2_PHOSPHO_SITE	PDOC00006
PS00006	559->563	CK2_PHOSPHO_SITE	PDOC00006
PS00008	164->170	MYRISTYL	PDOC00008
PS00008	256->262	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00009	167->171	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22f21.3)

DKFZphfbr2_22h13

group: transmembrane protein

DKFZphfbr2_22h13 encodes a novel 520 amino acid protein, with similarity to *Drosophila melanogaster* EG:39E1.3.

The protein contains an ATP/GTP A Prosite pattern (P-loop). This loop interacts with one of the phosphate groups of a A or G nucleotide. It is found in numerous ATP- or GTP-binding proteins, such as ATP synthase alpha and beta subunits, Myosin heavy chains, Kinesin heavy chains and kinesin-like proteins, Dynamins and dynamin-like proteins, several kinases, DNA and RNA helicases, GTP-binding elongation factors and the Ras family of GTP-binding proteins. Additionally, the novel protein contains one putative transmembran domain.

No informative BLAST results; no predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

AC004780_1, differences to predicted genmodel

membrane regions: 1

AC004780_1, differences to predicted genmodel

complete cDNA, complete cds, EST hits
on genomic level encoded by AC004780,
differences to predicted genmodel!
TRANSMEMBRANE 1

Sequenced by AGOWA

Locus: unknown

Insert length: 2292 bp

Poly A stretch at pos. 2272, polyadenylation signal at pos. 2255

```
1  GGGGGAGGGA  ACTGATCTCA  GCTCGGGCCC  GCGTTACATC  CTCCTCCTCT
51  TCTTCTCTTCG  GCCCAGCTTT  CCTTAGGGGC  TGCAACCCGG  ACGCCGAGGC
101  CGGTTTCGGA  GTGGGGAGTG  CCCATTTTCT  CTCCTTCCCA  CGTTCCTGGC
151  CCCCAGACGC  CATTTCAGG  CGGGTGGCTT  GGGTCAGCCT  CCCC GCCCCC
201  ACCCGACTCC  CGTCACGGGA  GAGCGCACAC  CGGCGCCCGA  GAACCAATCA
251  GCAGCCGCGT  TAGGTAACCA  TGTCTGAGTC  TGGACACAGT  CAGCCTGGAC
301  TCTATGGGAT  AGAGCGGCGG  CGACGGTGGA  AGGAGCCTGG  CTCGTGGTGGC
351  CCCCAGAATC  TCTCTGGGCC  TGGTGGTCCG  GAGAGGGACT  ACATTGCACC
401  ATGGGAAAGA  GAGAGAAGGG  ATGCCAGCGA  AGAGACAAGC  ACTTCCGTCA
451  TGCAGAAAC  CCCCATCATC  CTCTCAAAAC  CTCACAGAGA  CCGGTCAAAA
501  CAGCCACCAC  CTCCAACAGC  CCCTGCTGCC  CCGCCTGCTC  CAGCCCTCT
551  GGAGAAGCCC  ATCGTTCTCA  TGAAGCCACG  GGAGGAGGGG  AAGGGGCCCTG
601  TGGCCGTGAC  AGGTGCCTCT  ACCCCTGAGG  GCACCGCCCC  ACCACCCCTT
651  GCAGCCCTGT  CGCCACCCAA  GGGGGAGAAG  GAGGGGCAGA  GACCCACACA
701  GCCTGTGTAC  CAGATCCAGA  ACCGGGGCAT  GGGCACTGCC  GCACCAAGCAG
751  CCATGGACCC  TGTCTGGGT  CAGGCCAAAC  TACTGCCCCC  AGAGCGCATG
801  AAGCACAGCA  TCAAGTTGGT  GGATGACCAG  ATGAATTGGT  GTGACAGTGC
851  CATCGAGTAC  CTGTTGGATC  AGACTGATGT  GTTGGTGGTT  GGTGTCCTGG
901  GCCTCCAGGG  GACAGGCAAG  TCCATGGTCA  TGTCAATTGT  GTCAGCCAAC
951  ACTCCAGAGG  AGGACCAGAG  GACTTATGTT  TTCCGGGCCC  AGAGCGCTGA
1001  AATGAAGGAA  CGAGGGGGCA  ACCAGACCAG  TGGCATCGAC  TTCTTTATTA
1051  CCCAAGAACG  GATTGTTTT  CTGGACACAC  AGCCCATCCT  GAGCCCTTCT
1101  ATCCTAGACC  ATCTCATCAA  TAATGACCGC  AAATGCCTC  CAGAGTACAA
1151  CCTTCCCCAC  ACTTACGTTG  AAATGCAGTC  ACTCCAGATT  GCTGCCTTCC
1201  TTTTCACGGT  CTGCCATGTG  GTGATTGTTG  TCCAGGACTG  GTTCACAGAC
1251  CTCAGTCTCT  ACAGGTTCC  GCAGACAGCA  GAGATGGTGA  AGCCCTCCAC
1301  CCGATCCCCC  AGCCACGAGT  CCAGCAGCTC  ATCGGGCTCC  GATGAAGGCA
1351  CCGAGTACTA  CCCCCACCTA  GTCTTCTTGC  AGAACAAAGC  TCGCCGAGAG
1401  GACTTCTGTG  CTCGGAAGCT  GCGGCAGATG  CACCTGATGA  TTGACCAGCT
1451  CATGGGCCAC  TCCACCTG  GTTACAAGGG  AACTCTGTCC  ATGTTACAAT
1501  GCAATGTCTT  CCGGGGGCTT  CCACCTGACT  TCCTGGACTC  TGAGGTCAAC
1551  TTATTCTCTG  TACCCTTCAT  GGACAGTGAA  GCAGAGAGTG  AAAACCCACC
1601  AAGAGCAGGA  CCTGGTTCCA  GCCCACTCTT  CTCCTGTGCT  CCGGGGTATC
1651  GTGGCCACCC  CAGTTTCCAG  TCCTTGGTGA  GCAAGCTCCG  GAGCCAAGTG
1701  ATGTCCATGG  CCGGCCACAC  GCTGTCACAC  ACGATCCTCA  CCGAGAAGAA
1751  CTGGTTCCAC  TACGCTGCCC  GGATCTGGGA  TGGGGTGAGA  AAGTCTCTG
1801  CTCTGGCAGA  GTACAGCCGC  CTGCTGGCCT  GAGGCCAAGG  AGAGGAATGT
1851  CATGCAGGGG  ACCTCCTGGG  TCCGCACTGT  ACTGCGAGGG  AGCACAGATG
1901  TCCATCCCCC  GCTGGGGTGG  AGAGCGGCAG  CAGGCCTGAT  GGATGAGGGA
1951  TCGTGGCTTC  CCGGCCCAGA  GACATGAGGT  GTCCAGGGCC  AGGCCCCCA
```



```

2001 CCCTCAGTTG GGGCTGTTCC GGGGGTGA CT GTGAGCGATC CCACCCCAAA
2051 CCTGAGATGG GGTAGCCCGT CCTGTGTCCT CCACAGGGAC AAGCAGTGGG
2101 AGGAGTCTGA ATGGTCACCA GGAAGCCCGG GCTCCATCTT GACCTCCTTT
2151 TTCAGGGACA GGAGCAACAG GCCCCTCTTC CCTGACTCTA AGCCCTTCCC
2201 TGTAAGGTGA GGCAGGGTCT GGAGAGCTCT TTATTGGAAC AGATCTGGTG
2251 GTTCAAATAA ACACAGTCAT GCAAAAAAAA AAAAAAAA AA

```

BLAST Results

Entry AC004780 from database EMBL:
Homo sapiens chromosome 19, cosmid F17127, complete sequence.
Score = 2616, P = 0.0e+00, identities = 524/525
15 exons Bp 8031-31789

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 270 bp to 1829 bp; peptide length: 520
Category: similarity to unknown protein
Prosites motifs: ATP_GTP_A (211-219)

```

1 MSEGHSQPG LYGIERRRRW KEPGSGGPQN LSGPGGRERD YIAPWERERR
51 DASEETSTSV MQKTPILSK PPAERSKQPP PPTAPAAPPA PAPLEKPIVL
101 MKPREEGKGP VAVTGASTPE GTAPPPPAAP APPKGEKEGQ RPTQPVYQIQ
151 NRGMGTAAPA AMPDVVGQAK LLPPERMKHS IKLVDDQMNW CDSAIEYLLD
201 QTDVLVVGVL GLQGTGKSMV MSLLSANTPE EDQRTYVFRA QSAEMKERGG
251 NQTSGIDFFI TQERIVFLDT QPILSPSILD HLINNDRLKP PEYNLPHTYV
301 EMQSLQIAAF LFTVCHVVIV VQDWFTDLSL YRFLQTAEMV KPSTFSPSHE
351 SSSSSGSGDEG TEYYPHLVFL QNKARREDFC PRKLQRMHLM IDQLMAHSHL
401 RYKGTLSMLQ CNVFPGLPPD FLDSEVNLFL VPFMDEAES ENPPRAGPGS
451 SPFSLPLPGY RGHPFSFQSLV SKLRSQVMSM ARPQLSHTIL TEKNWFHYAA
501 RIWDGVRKSS ALAEYSRLLA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22h13, frame 3

TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19,
cosmid F17127, complete sequence., N = 2, Score = 1264, P = 1.3e-231

TREMBL:CEY54E2A_1 gene: "Y54E2A.2"; Caenorhabditis elegans cosmid
Y54E2A, N = 2, Score = 219, P = 1.4e-15

>TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence.
Length = 528

HSPs:

Score = 1264 (189.6 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
Identities = 254/302 (84%), Positives = 264/302 (87%)

```

Query: 46 ERERRDASEETSTSVMQKTPILSKPPAERSKQPPPTAPAAPAPAPLEKPIVLMPRE 105
      E+ER D+ + S +Q+T + R + P + A APLEKPIVLMPRE
Sbjct: 39 EKER-DSDSDFSP--LQQTGECQRRDKHFRHAENPHHPLKTSSRA-APLEKPIVLMPRE 94

Query: 106 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV 165
      EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV
Sbjct: 95 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV 154

Query: 166 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 225
      VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS
Sbjct: 155 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 214

```


Query:	226	ANTPEEDQRTYVVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN	285
		ANTPEEDQRTYVVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN	
Sbjct:	215	ANTPEEDQRTYVVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN	274
Query:	286	DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRFLQTAEMVKPSTP	345
		DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYR	K ++
Sbjct:	275	DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRLWDLGCKCKSNSH	334
Query:	346	SP 347	
		SP	
Sbjct:	335	SP 336	
Score = 993 (149.0 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231			
Identities = 189/189 (100%), Positives = 189/189 (100%)			
Query:	332	RFLQTAEMVKPSTPSPSHESSSSSGSDGTEYYPHLVLQNKARREDFCPRKLRQMHLMI	391
		RFLQTAEMVKPSTPSPSHESSSSSGSDGTEYYPHLVLQNKARREDFCPRKLRQMHLMI	
Sbjct:	340	RFLQTAEMVKPSTPSPSHESSSSSGSDGTEYYPHLVLQNKARREDFCPRKLRQMHLMI	399
Query:	392	DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFVLPFMDSEAESENPPRAGPGSS	451
		DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFVLPFMDSEAESENPPRAGPGSS	
Sbjct:	400	DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFVLPFMDSEAESENPPRAGPGSS	459
Query:	452	PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA	511
		PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA	
Sbjct:	460	PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA	519
Query:	512	LAEYSRLLA 520	
		LAEYSRLLA	
Sbjct:	520	LAEYSRLLA 528	

Pedant information for DKFZphfbr2_22h13, frame 3

Report for DKF2phfbr2_22h13.3

```
[LENGTH]           520
[MW]                57650.81
[pI]                6.52
[HOMOL]             TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence. 0.0
[PROSITE]           ATP_GTP_A             1
[PROSITE]           MYRISTYL              8
[PROSITE]           CAMP_PHOSPHO_SITE     1
[PROSITE]           CK2_PHOSPHO_SITE      8
[PROSITE]           GLYCOSAMINOGLYCAN    1
[PROSITE]           PKC_PHOSPHO_SITE      3
[PROSITE]           ASN_GLYCOSYLATION     2
[KW]                TRANSMEMBRANE         1
[KW]                LOW COMPLEXITY        11.73 %
```

[illegible]

```
SEQ      MQKTPIILSKPPAERSKQPPPTAPAAPAPAPLEKPIVLMPKEEGKGPAVAVTGASTPE
SEG      .....xxxxxxxxxxxxxxxxx.....
PRD      ecccccecccccccccccccccccccccccccccccceeeccccccccccceeeccccc
MEM
```

```
SEQ      GTATPPPPAAAPPKGEKEGRPTQPVYQIQNRGMGTAAAPAMPDVPVGAKLLPPERMKHS
SEG      . xxxxxxxxxxxx .....
PRD      cccccccccccccccccccccccceeeeccccccccccccceecceeccchhhhhh
MEM
```

```
SEQ      IKLVDDQMNWCDSAIEYLLDQTVLVVGVGLGQTGKSMVMSLLSANTPEEDQRTYVFRA
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      hhhhcccchhhhhhhhccccceeeeeeccccccchhhhhhhccccchhhhhheeee
MEM
```

```
SEQ      QSAEMKERGNGQTSGIDFFITQERIVFLDTQPILSPSILDHLINDRKLPPEYNLPHTYV
SEG
PRD      hhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccch
MEM
```

SEQ EMQSLQIAAF LFTVCHVVIVVQDWFTDLSLYRFLQTAEMVKPSTPSPSHESSSSSGSDEG
 SEGXXXXXXXXXXXXXXXXXXXXX.....


```

PRD      hhhhhhhhhhhhhhhheeeeeccchhhhhhhhhhhhhcccccccccccccccccc
MEM      MMMMMMMMMMMMMMMMMMMMM.....

SEQ      TEYYPHLVFLQNKARREDFCPKRLRQMHLMIDQLMAHSHLRYKGTLSMLQCNVFPGLPPD
SEG      .....
PRD      cccccceeehhhhhhccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc
MEM      .....

SEQ      FLDSEVNLFLVPFMDSEAESENPPRAGPGSSPLFSLLPGYRGHPSFQSLVSKLRSQVMSM
SEG      .....
PRD      chhhhhhheeeeeccccccccccccccccccccceccccccccchhhhhhhhhhhhhhh
MEM      .....

SEQ      ARPQLSHTILTEKNWFHYAARIWDGVRKSSALAEYSRLLA
SEG      .....
PRD      hhhhhhhheeeccchhhhhhhhhhhhhcchhhhhhhhhhhccc
MEM      .....

```

Prosites for DKFZphfbr2_22h13.3

PS00001	30->34	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00002	32->36	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	507->511	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	215->218	PKC_PHOSPHO_SITE	PDOC00005
PS00005	491->494	PKC_PHOSPHO_SITE	PDOC00005
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	254->258	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	436->440	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	211->217	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008
PS00008	505->511	MYRISTYL	PDOC00008
PS00017	211->219	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphfbr2_22h13.3)

DKFZphfbr2_22i4

group: brain derived

DKFZphfbr2_22i4.1 encodes a novel 228 amino acid protein with similarity to the N-terminus of human p52rIPK.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Human P52rIPK N-terminus

complete cDNA, complete cds, few EST hits
function of P52rIPK, repressor of p58IPK protein kinase inhibitor
upstream regulator of interferon induced proteins

Sequenced by AGOWA

Locus: unknown

Insert length: 4748 bp
Poly A stretch at pos. 4726, polyadenylation signal at pos. 4709

```
1 TGGGTCCGGT CCTAGGGTCA CACCCACCGC AGGGTCTGGC TTGGTACAGT
51 TGGGTGCATG CAGAAGTAGG TGGAGCTGCT GTTGCAGCCT TGAGAGAGTT
101 TTATTGTAAA ACTCTTGTA TTTATAGTAA TCGGAGGGGA AAACACCTCT
151 TCCTTTTAAT TGCTCTGAGG ACCGCTGCCA AAGAAACGCA GTAGATCCGC
201 TCCTCTTTGG GGGCGGGGAG AAAGAACGGG TTGTGTCCGC CATGTTGGTG
251 AAGTCAAGCG AAGGCGACTA GAGCTCCAGG AGGGCCAGTT CTGTGGGCTC
301 TAGTCGGCCA TATTAATAAA GAGAAAGGGA AGGCTGACCG TCCTTCGCCT
351 CCGCCCCCAC ATACACACCC CTTCTTCCCA CTCCGCTCTC ACGACTAAGC
401 TCTCACGATT AAGGCACGCC TGCCTCGATT GTCCAGCCTC TGCCAGAAGA
451 AAGCTTAGCA GCCAGCGCCT CAGTAGAGAC CTAAGGGCGC TGAATGAGTG
501 GGAAAGGGAA ATGCCGACCA ATTGCGCTGC GCGGGGCTGT GCCACTACCT
551 ACAACAAGCA CATTAACATC AGCTTCCACA GGTTCCTTTT GGATCCTAAA
601 AGAAGAAAAG AATGGGTTCC CTTGGTTAGG CGCAAAAATT TTGTGCCAGG
651 AAAACACACT TTTCTTTGTT CAAAGCACTT TGAAGCCTCC TGTTTTGACC
701 TAACAGGACA AACTCGACGA CTTAAATGAG ATGCTGTGCC AACCATTTTT
751 GATTTTTGTA CCAATATAAA GTCTATGAAA CTCAGTCAA GGAATCTTTT
801 GAAGAAAAAC AACAGTTGTT CTCCAGCTGG ACCATCTAAT TTAATCAAA
851 ACATTAGTAG TCAGCAAGTA CTACTTGAAC ACAGCTATGC CTTTAGGAAT
901 CCTATGGAGG CAAAAAAGAG GATCATTAAA CTGGAAAAAG AAATAGCAAG
951 CTTAAGAAGA AAAATGAAAA CTTGCCTACA AAAGGAACGC AGAGCACTC
1001 GAAGATGGAT CAAAGCCACG TGTTTGGTAA AGAATTTAGA AGCAATAGT
1051 GTATTACCTA AAGGTACATC AGAACACATG TTACCAACTG CCTTAAGCAG
1101 TCTTCCCTTG GAAGATTTA AGATCCTTGA ACAAGATCAA CAAGATAAAA
1151 CACTGCTAAG TCTAAATCTA AAACAGACCA AGAGTACCTT CATTTAAATT
1201 TAGCTTGCAC AGAGCTTGAT GCCTATCCTT CATTCTTTTC AGAAGTAAAG
1251 ATAATTATGG CACTTATGCC AAAATTCATT ATTTAATAAA GTTTTACTTG
1301 AAGTAACATT ACTGAATTG TGAAGACTTG ATTACAAAAG AATAAAAAAC
1351 TTCAATGGA AATTTTATTT GAAAATGAGT GGAAGTGCCT TACATTAGAA
1401 TTACGGACTT AAAAATTTTG CTAATAAATT GTGTGTTTGA AAGGTGTTTT
1451 TTGTTTTTGT CTTTTTAAAC TACTGTTAAA AGAACAGCTT ATGATAAGTA
1501 ATATGTTTAA CTTAGAGAAG AATTTTTTCC TGTACCAAAG TTGGCATATT
1551 GCATTCTAAA TAAGATGCTA AATAAGAGTT AACCACATT CAACATGACC
1601 TTA AAACTGC TGGGTTTTGT ATTAATTAAA TTATAATTGG CACTGTGATT
1651 TGA AAAATTT ATAGAAAAAA AGGTACAGGG CAAGTTTTTA AATTAAACT
1701 TTTCTATATT TGTTTTACCA GTAAAAGTGA GCTTATCATG GCCTCTCTCA
1751 TAAGAATGAT TTTAAAATAG GTTGTAATAA ATTTTGAAAA TATTGAAATG
1801 TGAAGTACCA TTCAGTCATC CAACTAGGT AAGGCCTCAA GTACTTTAAA
1851 CTAGTAAAAT CTAGTAGCTG ATAATATTCA CCTAAGTAAG TGTGTAAAAA
1901 TAATTCAGAG TTCAGGACCT AGCTTAGATA AATGTATACT ACTCTTTTTC
1951 TCATAGTAAA AATCTTACAT TTCCAACCTC AAAATTTGGT GTTCCATATT
2001 TGTTGATAAC CAAAACCTCT AAGGTTTTTT GTTTTCTTTT TAACTACTTT
2051 GCAAATGCAT ACTATACCTC AGAAATAGTG TATCAATATA GTGGGCTTTT
2101 TTTTCTCTCT TCATAAACC ACAGTAAAT TTAATCACAG GAACTACTTT
2151 ATATCTTAC ACTTTGTATT GATAACTTAA AATGGCATCA GTTTATCTTA
2201 GACATCAGCT TGCTTTTAT CTCCTTTTTT AGTGAGTGAA ATAGAGCAAC
2251 TAGCATGCCT GTGTCCAG CTAAGTGGGA GGCTAAGGTG GGAAGATCAA
2301 TTGAACCTAG GAGGTTGAGG CTATAGTGAG CTGTGATTGC ACGACTGCAC
2351 TCCAGCCTGG GCAATGGAGT GAGACTCCTG TCTCTAAAAC AGCAACAACA
2401 AAAATAAAGC AACCATAGTG CATAAGGGAA ATTAATGTT CCCTATAGAA
2451 ATATGTGTAT GTCGTGATA GTGGTATGCA AATGCTAATT ATTTTATAAA
2501 ATAAAAAGTTC AGAATATTC TTATCATTGC CACTTGAACA ATTAAGGGT
2551 TTGCTTTATT TCACTAATGT TTAATAGGAA CCCTTTGCTT CAAACAGCTT
```



```

2601 TGTGAAATC ATGTAAAAAT TGTGAATAG AGAATCAAGT TATTTAACTC
2651 AACTTATTTA ATTCAAGCTT GTGATACTAA CATACAAAGG TAGCATAAAC
2701 CAAGTCATAA ATTGCTGTAA TCTTTCCTGT AGAGTAATAG CTACTTCATG
2751 ATTTTAAAAA AAATTTTCATT TTTTGTCTAT TTAGGATTGC ATTTGCTTGG
2801 CTCCTAGTAA CAATTCTTTT ACAGTATTAG CACTCTCTTT ACTAAGGAAT
2851 GCCTCCCAAG GAAATGCAAA GGTAGGAAAA GTCTCTTAGA ATGCCCATGA
2901 GGTATTTAAA ACAGATATTT ATGAAAATCT TTTTGTGAAT GTTATAAATC
2951 TTGCTAGTTA TTTTATCTTT ATCTTAAGTA TTAGATGTAG TTCCTTGGAA
3001 TTGTCATTAC ATATTATTT TTTTCTAGTG TGGTTTCAAA TAACTTTTGG
3051 CCAACATATA ATCATCATCA AACATTCACT GACCATATCT ATTTTATAAC
3101 TCAAAATAAG TTGGACAAAT AATCATTTTA ATAAAACTA TTTTTCCAA
3151 GTATAACAC TGTCATGTGG TTCACCTTC ACCCCAGATA CAAACACTT
3201 ATTTGTGTAG CCCAGTTCCC ATCTACAGTA ATACCTTGAA ACCTTAATAA
3251 ATTTTAAAAA TCATAAAAAA AAAATATTGT AAAATACAAC AAATTTTGGG
3301 CAAGGTTACT TCATCTTCAT TCATTATTAC CTGACAGTAT TAAACTACTA
3351 CTCATAAATT TTAGAGTAAA CTTTCTGTG TTTTCCCCGT GATTTTCATT
3401 GTGCTGTCTT GACAACATGC TCCAAACTCT TTGCATCAA TTGTTTATT
3451 AACATACATT TGTCTACCTT AAAACTAGCT TTATTCACAG AGAAAGACCT
3501 AAAAGGAGTC TATTAATAATG CTGCTTCAG TTGTAGATT TTTTATTAAC
3551 TCACTCTGAC CATAAACTAA CTGAAATTAT AATGGATTTT TTTTCTCTC
3601 CCGGTCACAA CACAGATCTT CTGTTCATTT GTTCTCTGTC TACTGGGCAC
3651 CAACCTCTAC AAAGAACCAG CCAAGGCTA GGTACTTGAT ATAAAAAGGA
3701 ATATTACATT ATTTTCTGCC CTCAAGTTGC TCTATCTCCT GAAAGAAACA
3751 AGTAATATTT ATAATACAAT ATGATAAATG CTACAAAAGA AATAGCTGTA
3801 AAGTCCTTTG GTAATGCTG TTGAATTGGA ATTCAGTAAG AACTATAAAC
3851 TGTAGACCTT TTTATAATCA AATGCTTTT TCTTGAAACA AAACAGATT
3901 CTCCTTATAT TGACTTAGCA AAGGAGGTAC AAGGACATTG GCATTTGACC
3951 TGAATTATGG TGTTTTATTG AATGAGCTAT AAGACAACAT TTTTACCCTT
4001 TAAAAATGAC ACTGAACAAA TGTGTTAATG GTATCTTTGT TAAAAGGAAA
4051 ACATAGCTAT AAATAAAATA CTACATCGAA ATCCAGCACT GGAGTTTCATT
4101 TGAATTTGA TATTTTGTGT AAAGTAACAA ACCTATTAAC ACAGATTTT
4151 AAAATAACTC AGAATCGTAT AAAGCACTT GGTACTTATT TGTCTCTTT
4201 TCCCTTACAT TCTGTGTGGT AGGTGGTATT ATCTCTGATT TACACATGAA
4251 GACATCCTTG TTAATGCAAT TTATTTATTC ATTCGGGCAT TTACTGTGTG
4301 CCAACTTGCA AAAGGAATAG AAATGTCTGT GATCTAGATA GTTCTAGATT
4351 GAACATAGAT TTTCTGCCAA CAAATCCTCT CTGCTGTTCA CATTATCCTT
4401 TGTTTAACGT ATGAACCAGG TTAATAAAAT AGGATAAATC ATGTGTCTTA
4451 GAATATGAAA ATAGTAAGGT CTTTGAGGTC ACTTGATCTT CTCTAAGTAG
4501 ACTTTATAAT ATTGTGTTT ATCTCATTTT TCAATATTAG AATACGGGTA
4551 GATTTTAATT TTGCTATAAT ATAGGAAATG GTTCATCTT GTACCAAAAT
4601 ATTGCATTCT TCTGATATTT AGACAGTTGG AAACCTTCTA AAATTGAGGA
4651 TTTTGTAGTG TATACTAAAT AATTGCATAT TCAAAAAAAT GTATTCTGAG
4701 TATGGTGATA TTAAACATTT TTCCCCAAAA AAAAAAATAA AAAAAAATAA

```

BLAST Results

No BLAST result

Medline entries

98107671:
Regulation of interferon-induced protein kinase PKR:
modulation of P58IPK inhibitory function by a novel protein,
P52rIPK

Peptide information for frame 1

ORF from 511 bp to 1194 bp; peptide length: 228
Category: similarity to known protein

```

1 MPTNCAAAGC ATTYNKHINI SFHRFPLDPK RRKEWVRLVR RKNFVPGKHT
51 FLCCKHEFAS CFDLTGQTRR LKMDAVPTIF DFCTHIKSMK LKSRNLLKKN
101 NSCSPAGPSN LKSNISSQOV LLEHSYAFRN PMEAKKRIK LEKEIASLRR
151 KMKTCLOQER RATRRWIKAT CLVKNLEANS VLPKGTSEHM LPTALSSLPL
201 EDFKILEQDQ QDKTLLSLNL KQTKSTFI

```

BLASTP hits

Entry AF007393_1 from database TREMBL:
product: "P52rIPK"; Homo sapiens P52rIPK mRNA, complete cds.
Score = 166, P = 2.5e-11, identities = 40/106, positives = 56/106

DKFZphfbr2_22k3

group: brain derived

DKFZphfbr2_22k3 encodes a novel 538 amino acid protein with weak similarity to extensins.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to extensins

complete cDNA, complete cds, few EST hits
CpG Island in 5' UTR complete cDNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2775 bp

Poly A stretch at pos. 2755, polyadenylation signal at pos. 2718

```
1 GGGGCTGCCC GCGCGCTCCA CGGTGCAGAG CTCTAAGCGC GCGGGCTGGC
51 AGGCTGCGGC GCGTCAAGGT CAGCCTGGAG CTGGGTGGCG GCCTGCCCTGG
101 GGGCGGGGGA CCCTACTGGA GGGCCGGGCT GGGGCTCCC AGCGCCTCGG
151 CCATATTGAA TAGCTTCGAC TGGACCGTCT TTGTCTGCGA AGTCCTGTCC
201 CAAGTTCAG CCGCGTCCCT GGGGCTGGG GCAGGAAGAG TCGCTGGCAG
251 CCGCGCGGCC CCAACTTGA GCTGGGACAC CACGTTTCCA GCTTGGAGTG
301 GGCCTTGAGC CTTGGGACTG ACCTCGCCCC CGGCTCACGT AGGCATCCTG
351 GAAATTGATT CCCCCAAGTC CTTGGTGGGG GAGCCGGACT TGGTCAAGAC
401 TGTACTTGTG GCAGGCGAAG AGATTGGAGG CGTTTGGCTC GTCCTGGCT
451 AGGGAGGTGA GACTCTCCGG TCAGCGTTGC TGGAACTCCC CCCATCCAGT
501 CCCTCCCTCA AGACTAAGGG CTACAGTAGT TTGTTGGGGC TCATTGCCCC
551 CTCACCCAG ATATCACCCCT GGAGATCTTA AAGACTCTCG AGAAAAGCCA
601 CGTGGGGGGC TGGTTCCCT GGGGCTTCCT GCCGTCCCC GACTGCCTCA
651 TTCTTTGGAG CGTCCCGGAT GTCTGCAAAG ATGTGGATTT GGACGTCTCT
701 GTGGAAGCCC TAAAGCCCGT GGGGACATTT AAGAAGATCG GCAAGGTGTT
751 CCGCAAGGAG GAGGACTCCA CGGTGGGGAT GCTGCAGATC GGGGAGGACG
801 TCGACTATTT GCTCATCCCC CGGGAGGTCA GGTGCGTGG GGGCGTCTGG
851 AGAGTCATCT CTAAGCCCGC CACCAAGGAA GCAGAATTTG GGGAGCGGCT
901 GACCCAGTTC CTGGAAGAAG AGGGCCGCAC CCTGGAGGAC GTGCCCCGCA
951 TCATGGAGAA GAGCACCCCG CACCCGCCCC AGCCCCCAA AAAGCCCCAG
1001 GAGCCCCGAG TGAGGAGGAG AGTGCAGCAG ATGGTGACTC CTCGCCCCCG
1051 GCTGGTCTGT GGCACGTACG ACAGCAGCAA CGCCAGCGAC AGCGAGTTCA
1101 GCGACTTCGA GACCTCCAGA GACAAGAGCC GCCAGGGCCC GCGCGGGGCG
1151 AAGAAGGTGC GCAAAATGCC CGTCAGCTAC CTGGGCAGCA AGTTCCTGGG
1201 AAGCGACCTG GAGAGTGAGG ATGATGAGGA ACTGGTCTGAG GCCTTCCTCC
1251 GCGCAGAGGA GAAGCAGCCC AGCGCGCCCG CTGCCCCCGG CCGCGTCAAC
1301 CTGCCAGTGC CCATGTTTGA GGACAACCTG GGGCCTCAGC TGTCCAAAGC
1351 GGACAGGTGG CCGGAGTATG TCAGCCAGGT GTCTTGGGGG AAGCTGAAGC
1401 GGAGGGTGAA GGGTTGGGCG CCGAGGGCGG GCCCGGGGT GGGCAGGCCC
1451 CGGCTGGCCT CCACCGCAGT GGAGAGCGCA GGGGTATCAT CGGCGCCAGA
1501 GGGCACCAGC CCGGGGGATC GCTTGGGAAA CGCGGGAGAT GTTTGTGTGC
1551 CCCAGGCTTC CCCTAGGCGA TGGAGGCCCA AGATCAACTG GGCCTCCTTT
1601 CCGCGCCGCA GGAAGGAGCA GACAGCACC ACAGGTCAGG GGGCAGACAT
1651 CGAGGCTGAT CAGGGGGGAG AGGCTGCAGA TAGTCAAAGG GAAGAGGCCA
1701 TAGCTGACCA GCGGGAAGGG GCTGCAGGTA ATCAGAGGGC TGGGGCCCCA
1751 GCTGACCAGG GGGCAGAGGC TGCAGATAAT CAGAGGGAAG AGGCTGCAGA
1801 TAATCAGAGG GCAGGGGGCC CAGCTGAGGA GGGGGCAGAG GCTGCAGATA
1851 ACCAGAGGGA AGAGGCTGCA GATAATCAGA GGGCAGAGGC CCCAGCTGAC
1901 CAGAGGTCAC AGGGCACAGA TAACCACAGG GAAGAGGCTG CAGATAATCA
1951 GAGGGCGGAG GCCCCAGCTG ACCAGGGGTC AGAGGTTACA GATAATCAAA
2001 GGGAAAGAGG CGTACATGAC CAGAGGGAAA GGGCCCCAGC TGTCCAGGGT
2051 GCAGATAATC AGAGGGGACA GGGCCGGGCT GGCAGAGGG CAGAGGCTGC
2101 ACATAATCAG AGGGCAGGGG CCCCAGGTAT CCAGGAAGCT GAAGTCTCAG
2151 CTGCCCAAGG GACCCACAGG ACAGCTCCAG GAGCCAGGGC CCGGAAACAG
2201 GTCAAGACAG TGAGGTTCCA GACCCCTGGA CGCTTTTCGT GGTTTTGCAA
2251 GCGCCCGAGA GCCTTCTGGC ACACTCCCG GTTGCCAACC CTGCCCAAGA
2301 GAGTCCCCAG GGCAGGAGAG GTCAGGAACC TCAGGGTGCT GAGGGCCGAG
2351 GCCAGAGCAG AAGCTGAGCA GGGAGAGCAA GAAGACCAGC TGTGAGGTGA
2401 GGGCTAGAGA CAGCCACGG GCCCTCCCTC CAAGTGTGGG AGGGAGAGAT
2451 GCTCTGCCTC TGAACCTCAA AGTGGAGGTG GAGTGCTGGC CACGTCTCCA
2501 CCTAACAAAC CTCTTTATTC TCTTGTAAAA GTTTTGTTC TGTCTTGATT
2551 TTTTAAAAA TTTTATAGAG ACAGGGTCTC ACTCTGTTGC CCAGGCTGGA
2601 GTGCAGTGGC ATGATCATAA CTCAGTCAG CCTCAAATT CTGGCCTCAA
2651 GTGATCTCC TGCCTCGGCC TCCCAAAATG CTGGGATTAC AGATGTGAGC
```


2701 CACCACACAC ACCATCTGAT TAAAAAATAA AAATACTGAT TCCCTGTAGC
 2751 AACCCAAAAA AAAAAAAAAA AAAAA

BLAST Results -----

Entry HS164A7F from database EMBL:
 H.sapiens CpG island DNA genomic MseI fragment, clone 164a7, forward
 read cpg164a7.ft1a .
 Score = 740, P = 3.0e-25, identities = 150/151

Medline entries -----

No Medline entry

Peptide information for frame 2 -----

ORF from 779 bp to 2392 bp; peptide length: 538
 Category: similarity to known protein

1 MLQIGEDVDY LLIPREVRLA GGWVRVSKP ATKEAEFRER LTQFLEEEGR
 51 TLEDVARIME KSTPHPPQPP KKPKEPRVRR RVQQMVTPPP RLVVGTYDSS
 101 NASDSEFSD FTSRDKSRQG PRRGKKVRKM PVSYLGSKFL GSDLESEDDE
 151 ELVEAFLRRQ EKQPSAPPAR RRVNLPVPMF EDNLGPQLSK ADRWREYVSQ
 201 VSWGKLRKRRV KGWAPRAGPG VGEARLASTA VESAGVSSAP EGTSPGDRLG
 251 NAGDVCVPQA SPRRWRPKIN WASFRRRRKE QTAPTQGCAD IEADQGGEEA
 301 DSQREEAIAD QREGAAGNQR AGAPADQGAE AADNQREEAA DNQRAGAPAE
 351 EGAEAADNQR EEAADNQRAE APADQRSQGT DNHREEAADN QRAEAPADQG
 401 SEVTDNQREE AVHDQREERAP AVQGADNQRA QARAGQRAEA AHNQRAGAPG
 451 IQEAEVSAAQ GTTGTAAPGAR ARKQVKTVERF QTPGRFSWFC KRRRAFVHTP
 501 RLPTLPKRVP RAGEVRNLRV LRAEARAEAE QGEQEDQL

BLASTP hits

Entry RNU67136.1 from database TREMBL:
 "A-kinase anchoring protein AKAP150"; Rattus norvegicus
 A-kinase anchoring protein AKAP150 mRNA, complete cds. Rattus
 norvegicus (Norway rat)
 Length = 714
 Score = 182 (64.1 bits), Expect = 1.2e-10, P = 1.2e-10
 Identities = 73/257 (28%), Positives = 104/257 (40%)

Alert BLASTP hits for DKF2phfbr2_22k3, frame 2

TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds., N = 1, Score = 178, P = 3.7e-11

>TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds.
 Length = 285

HSPs:

Score = 178 (26.7 bits), Expect = 3.7e-11, P = 3.7e-11
 Identities = 60/217 (27%), Positives = 97/217 (44%)

Query: 269 INWASFRRRRKEQTAPTQGA-DIEADQGGEEAADSQRE-EAIAQDQ---REGAAGNQRA 323
 +N + + + E G+G D E E +D+ E E I Q E A N+ AG+
 Sbjct: 47 LNKNGKGNKYEDLQEEGEGENDDEHSNSESDNDEENEIIVGQDGSNEKAGSNEEAGS 106
 Query: 324 PADQGAEEAADNQREEAADNQRAEAPAEAGA--EAADNQRE---EEAADNQRAEAPADQRS 377
 G+ E+A N++AG+ E G+ EA N+ EEA N++A + S
 Sbjct: 107 NEKAGSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGS 166
 Query: 378 QGTDNHREEAADNQRAEAPADQGSSEVTDNQREEAVHDQREERAPAVQGADNQRAQAR---AG 435
 EEA N++A + + GS E+A +++ + G+ N++A + AG
 Sbjct: 167 NEKAGSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEEAGS-NEKAGSNEEAG 225
 Query: 436 QRAEAAHNQRAA---PGIQEAEVSAAQGTGTGA-PGA 469

Report for DKFZphfbr2_22k3.2

```
[LENGTH]          538
[MW]               59402.19
[pI]              8.72
[HOMOL]           TREMBL:AF037364_1 gene: "MA1"; product: "paraneoplastic neuronal antigen MA1";
Homo sapiens paraneoplastic neuronal antigen MA1 (MA1) mRNA, complete cds. 4e-10
[PROSITE]         AMIDATION          1
[PROSITE]         MYRISTYL           12
[PROSITE]         CK2_PHOSPHO_SITE    11
[PROSITE]         PKC_PHOSPHO_SITE    6
[PROSITE]         ASN_GLYCOSYLATION   1
[KW]              All_Alpha
[KW]              LOW COMPLEXITY      18.03 %
```

```

SEQ      MLQIGEDVDYLLIPREVRLAGGVWRVISKPATKEAEFRERLTQFLEEGRTELDVARIME
SEG
PRD      cccccccccccccccccccccceeeeeccchhhhhhhhhhhhhccchhhhhhhhh

SEQ      KSTPHPPQPPKKPKPRVRRRVQQMTPPRLVVGTYDSSNASDSEFSDFETSRDKSRQG
SEG      . . . . xxxxxxxxxxxxxxxxxxxxxxx . . . .
PRD      hccccccccccccccccchhhhhhhhhccccceeeeecccccccccccccccccccccc

SEQ      PRRGKKVRKMPVSYLGSKFLGSDLESEDEELVEAFLRRQEKQPSAPPARRRVNLPVPMF
SEG      . . . . . xxxxxxxxxxxxxx . . . . .
PRD      cccccccccceeeccccccccccccchhhhhhhhhhhhhhhcccccchhhhhcccccc

```



```

SEQ  EDNLGPQLSKADRWREYVSQVSWGKLRVRKGWAPRAGPGVGEARLASTAVESAGVSSAP
SEG  .....
PRD  cccccccchhhhhhhhhheeeccchhhhhccccccccchhhhhhhhhcccccc

SEQ  EGTSPGDRNLGNAGDVCVPQASPRRWRPKINWASFRRRRKEQTAPTGGADIEADQGGEAA
SEG  .....
PRD  cccccccccccccceeeccccccccccccchhhhhhhhhhhccccchhhhhccchhh

SEQ  DSQREEAIADQREGAAGNQRAGAPADQGAEAADNQREEAADNQAGAPAEEGAEEAADNQ
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhh

SEQ  EEAADNQRAEAPADQRSQGTDNHREEAADNQRAEAPADQGSSEVTDNQREEAVHDQREAP
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  AVQGADNQRAQARAGQRAEAAHNQRAGAPGIQEAESAAQGTGTAPGARARKQVKTFRF
SEG  .....
PRD  hhccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhccccccccchhhhhhhhhhh

SEQ  QTPGRFSWFCKRRRAFHWHTPRLPTLPKRVPRAGEVRNLRVLAERAEAEQGEQEDQL
SEG  .....
PRD  cccccceehhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcc

```

Prosites for DKFZphfbr2_22k3.2

PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	273->276	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	499->502	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	112->116	CK2_PHOSPHO_SITE	PDOC00006
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008
PS00008	242->248	MYRISTYL	PDOC00008
PS00008	296->302	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	328->334	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	461->467	MYRISTYL	PDOC00008
PS00008	464->470	MYRISTYL	PDOC00008
PS00009	123->127	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22k3.2)

DKFZphfbr2_22k8

group: brain derived

DKFZphfbr2_22k8 encodes a novel 172 amino acid protein without similarity to known proteins.

No informative BLAST results; no predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: /map="7"

Insert length: 2789 bp

Poly A stretch at pos. 2769, polyadenylation signal at pos. 2756

```
1  GGGGGAGCCA TGAGGCGCCA GCCTGCGAAG GTGGCGGCGC TGCTGCTCGG
51  GCTGCTCTTG GAGTGCACAG AAGCCAAAAA GCATTGCTGG TATTTGGAAG
101 GACTCTATCC AACCTATTAT ATATGCCGCT CCTACGAGGA CTGCTGTGGC
151 TCCAGGTGCT GTGTGCGGGC CCTCTCCATA CAGAGGCTGT GGTACTTCTG
201 GTTCCTTCTG ATGATGGGCG TGCTTTTCTG CTGCGGAGCC GGCTTCTTCA
251 TCCGGAGGCG CATGTACCCC CGGCCGCTGA TCGAGGAGCC AGCCTTCAAT
301 GTGTCTTACA CCAGGCAGCC CCCAAATCCC GGCCAGGAG CCCAGCAGCC
351 GGGGCCGCCC TATTACACTG ACCCAGGAGG ACCGGGGATG AACCCGTGCG
401 GGAATTCACG GGCATGGCT TTCCAGGTCC CACCAACTC ACCCCAGGGG
451 AGTGTGGCCT GCCCGCCCCC TCCAGCCTAC TGCAACACGC CTCGCCCCCC
501 GTACGAACAG GTAGTGAAGG CCAAGTAGTG GGGTGCCAC GTGCAAGAGG
551 AGAGACAGGA GAGGGCCTTT CCCTGGCCTT TCTGTCTTCG TTGATGTTCA
601 CTTCCAGGAA CGGTCTCGTG GGCTGCTAAG GGCAGTTCCT CTGATATCCT
651 CACAGCAAGC ACAGCTCTCT TTCAGGCTTT CCATGGAGTA CAATATATGA
701 ACTCACACTT TGTCTCCTCT GTTGCTTCTG TTCTGACGC AGTCTGTGCT
751 CTCACATGGT AGTGTGGTGA CAGTCCCCGA GGGCTGACGT CCTTACGGTG
801 GCGTGACCAG ATCTACAGGA GAGAGACTGA GAGGAAGAAG GCAGTGCTGG
851 AGGTGCAGGT GGCATGTAGA GGGGCCAGGC CGAGCATCCC AGGCAAGCAT
901 CCTTCTGCCC GGGTATTAAT AGGAAGCCCC ATGCCGGGCG GCTCAGCCGA
951 TGAAGCAGCA GCCGACTGAG CTGAGCCCAG CAGGTCATCT GCTCCAGCCT
1001 GTCCCTCTCGT CAGCCTTCCT CTTCAGAAG CTGTTGGAGA GACATTCAGG
1051 AGAGAGCAAG CCCCTTGTC TGTCTTCTG TCTGTTTATA TCCTAAAGAT
1101 AGACTTCTCC TGCACCGCCA GGGAAAGGATA GCACGTGCAG CTCTCACC GC
1151 AGGATGGGGC CTAGAATCAG GCTTGCCTTG GAGGCTGAC AGTGATCTGA
1201 CATCCACTAA GCAAATTTAT TTAAATTCAT GGGAAATCAC TTCCTGCCCC
1251 AAAGTGAAGC ATTGCATTT GTGAGCTCTT GGTCTGATTT GGAGAAAGGA
1301 CTGTTACCCA TTTTCTTGGT GTGTTATGGA AAGTGCATGT AGAGCGTCTC
1351 GCCCTTTGAA ATCAGACTGG GTGTGTGTCT TCCCTGGACA TCACTGCCTC
1401 TCCAGGGCAT TCTCAGGCC GGGGGTCTCC TTCCTCAGG CAGCTCCAGT
1451 GTTGGGTTCT GAAGGGTGCT TCAAAAACGG GGCACATCTG GCCGGGAAGT
1501 CACATGGACT CTTCCAGGGA GAGAGACCAG CTGAGGCGTC TCTCTGTAG
1551 GTTGTGTTGG GTCTAAGCGG GTGTGTGCTG GGCTCCAAGG AGGAGGAGCT
1601 TGCTGGGAAA AGACAGGAGA AGTACTGACT CAACTGCACT GACCATGTTG
1651 TCATAATTAG AATAAAGAAG AAGTGGTCGG AAATGCACAT TCCTGGATAG
1701 GAATCACAGC TCACCCAGG ATCTCACAGG TAGTCTCTG AGTAGTTGAC
1751 GGCTAGCGGG GAGCTAGTTC CGCCGCATAG TTATAGTGTT GATGTGTGAA
1801 CGCTGACCTG TCCTGTGTGC TAAGAGCTAT GCAGCTTAGC TGAGGCGCCT
1851 AGATTACTAG ATGTGCTGTA TCACGGGGAA TGAGGTGGGG GTGCTTATTT
1901 TTTAATGAAC TAATCAGAGC CTCTTGAGAA ATTGTTACTC ATTGAACTGG
1951 AGCATCAAGA CATCTCATGG AAGTGGATAC GGAGTGATT GGTGTCCATG
2001 CTTTTCACCT TGAGGACATT TAATCGGAGA ACCTCCTGGG GAATTTTGTG
2051 GGAGACACTT GGGAAACAAA CAGACACCCT GGAATGCAG TTGCAAGCAC
2101 AGATGCTGCC ACCAGTGTCT CTGACCACCC TGGTGTGACT GCTGACTGCC
2151 AGCGTGGTAC CTCCCATGCT GCAGGCCTCC ATCTAAATGA GACAACAAAG
2201 CACAAATGTT ACTGTTTACA ACCAAGACAA CTGCGTGGGT CCAAACACTC
2251 CTCTTCTCTC AGGTCAATTG TTTTGCATTT TTAATGTCTT TATTTTGTG
2301 AATGAAGAAC CACACTAAGC TGCCCTGGA ATCGGGTGCA GCTGAATAGG
2351 CACCCAAAAG TCCGTGACTA AATTCCGTTT GTCTTTTGA TAGCAAATTA
2401 TGTTAAGAGA CAGTGATGGC TAGGGCTCAA CAATTTTGTG TTCCCATGTT
2451 TGTGTGAGAC AGAGTTTGTG TTCCCTTGAA CTGGTTAGA ATTGTGCTAC
2501 TGTGAACGCT GATCCTGCAT ATGGAAGTCC CACTTTGGTG ACATTTCTCTG
2551 GCCATTCTTG TTTCCATTGT GTGGATGGTG GGTGTGCCCC ACTTCTCTGGA
2601 GTGAGACAGC TCCTGGTGTG TAGAATTCCT GGAGCGTCCG TGGTTCAGAG
2651 TAACTTGAA GCAGATCTGT GCATGCTTTT CCTCTGCAGC AATTGGCTCG
2701 TTTCTCTTTT TTGTTCTCTT TTGATAGGAT CCTGTTTCCT ATGTGTGCAA
```


2751 AATAAAATA AATTGGGCA AAAAAAAAAA AAAAAAAAAA

BLAST Results

Entry HS671255 from database EMBL:
human STS SHGC-11828.
Length = 400
Minus Strand HSPs:
Score = 1822 (273.4 bits), Expect = 4.8e-76, P = 4.8e-76
Identities = 382/397 (96%), Positives = 382/397 (96%),

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 525 bp; peptide length: 172
Category: putative protein
Classification: unset

1 MRRQPAKVAA LLLGLLLECT EAKKHCWYFE GLYPTYIYICR SYEDCCGSRC
51 CVRALSIQRL WYFWFLMMG VLFCCGAGFF IRRRMYPPL IEEPAFNVS
101 TRQPPNPGPG AQQPGPPYYT DPGGPGMNPV GNSTAMAFQV PPNSPQGSVA
151 CPPPPAYCNT PPPPYEQVVK AK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22k8, frame 1

PIR:S14970 extensin class I (clone w17-1) - tomato, N = 1, Score = 118,
P = 2.3e-07

>PIR:S14970 extensin class I (clone w17-1) - tomato
Length = 132

HSPs:

Score = 118 (17.7 bits), Expect = 2.3e-07, P = 2.3e-07
Identities = 30/82 (36%), Positives = 35/82 (42%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
PPP P Y + PP P P P P YY P P +P + P SP
Sbjct: 32 PPSPPSPPP--PYYYKSPPPPSPSP--PPPYYYKSPPPPDPSPPPPYYYKSPPPPSPSP 87

Query: 147 GSVACPPPPAYCNTPPPP--YEQV 168
PPPP Y + PPPP YE +
Sbjct: 88 PPSPPSPPPPTYSSPPPPPPFYENI 111

Score = 104 (15.6 bits), Expect = 6.9e-06, P = 6.9e-06
Identities = 28/78 (35%), Positives = 34/78 (43%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
PP P + Y + PP P P P P YY P P +P ++ PP P
Sbjct: 1 PPSPPSPPPY---YYKSPPPPSPSP--PPPYYYKSPPPPSPSP---PPPYYYKSPP-PPS 51

Query: 147 GSVACPPPPAYCNTPPPP 164
S PPPP Y +PPPP
Sbjct: 52 PS---PPPPYYKSPPPP 66

Score = 102 (15.3 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 30/78 (38%), Positives = 33/78 (42%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
PPP P Y + PP P P P P YY P P +P S + PP P
Sbjct: 48 PPSPPSPPP--PYYYKSPPPPDPSP--PPPYYYKSPPPPSPSPPPPPSPS-----PP-PPT 97


```

Query:      147 GSVACPPPPPAYCNTPPPP 164
           S PPPP Y N P PP
Sbjct:     98 YSSPPPPPPFYENIPLPP 115

Score = 95 (14.3 bits), Expect = 2.4e-04, P = 2.4e-04
Identities = 24/61 (39%), Positives = 29/61 (47%)

Query:      104 PPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQGSVACPPPPPAYCNTPPPP 163
           PP+P P P P YY P P +P ++ PP P S PPPP Y +PPP
Sbjct:      1 PPSPS-----PPPYYYKSPPPSPSP---PPPYYYKSPP-PPSPS---PPPPYYKSPPP 49

Query:      164 P 164
           P
Sbjct:      50 P 50

Score = 68 (10.2 bits), Expect = 4.2e+00, P = 9.8e-01
Identities = 24/69 (34%), Positives = 29/69 (42%)

Query:      87 PPPLIEEPAFNVSYTRQPP---NPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPN 133
           PPP P Y PP +P P + P PP Y+ P P P + + PP
Sbjct:      63 PPPDPSPPPPPYYKSPPPSPSPPPSPSPPPPYSSPPPPP--PFYENIPL----PPV 116

Query:      144 SPQGSVACPPPP 155
           S A PPPP
Sbjct:      117 IGV-SYASPPPP 127

```

Peptide information for frame 3

ORF from 0 bp to 368 bp; peptide length: 123
Category: questionable ORF
Classification: unset

1 GSHEAPACEG GGAAARAALG VHRSQKALLV FRRTLNNLLY MPLLRGLLWL
51 QVLCAGPLHT EAVVLLVPSD DGRAFLRLSR LLHPEAHVPP AADRGASLQC
101 VLHQAAPKSR PRSPAAGAAL LH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22k8, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22k8, frame 1

Report for DKFZphfbr2_22k8.1

```

[LENGTH]      172
[MW]           19194.47
[pI]           8.77
[KW]           SIGNAL_PEPTIDE 23
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 27.33 %

SEQ MRRQPAKVAALLGLLLECTEAKKHCWYFEGLYPTYIICRSYEDCCGSRCCVRALSIQRL
SEG ..... xxxxxxxx .....
PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccceeeccccccccccchhhhhhhhhhh
MEM .....

SEQ WYFWFLLMGVLFCCGAGGFIRRRMYPPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYT
SEG ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccecccccccccccccccccc
MEM ..... MMMMMMMMMMMMMMMM .....

SEQ DPGGPGMNPVGNSTAMAFQVPNPSQGSVACPFPFPAYCNTPPPPYEQVVKAK
SEG xxxxxx ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD ccccccccccccccccccecccccccccccccccccccccccccccccccccccccc
MEM .....

```

(No Prosite data available for DKFZphfbr2_22k8.1)

(No Pfam data available for DKFZphfbr2_22k8.1)

Pedant information for DKFZphfbr2_22k8, frame 3

Report for DKFZphfbr2_22k8.3

[LENGTH] 122
[MW] 12854.08
[pI] 10.27
[KW] All_Alpha
[KW] LOW_COMPLEXITY 25.41 %

SEQ GSHEAPACEGGGAAARAALGVHRSQKALLVFRRTLSNLLYMPLLRGLLWLQVLCAGPLHT
SEGXX
PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhccccccccchhhhhhhcccccc

SEQ EAVVLLVPSDDGRAFLRLHPEAHVPPAADRGASLQCVLHQAAPKSRPRSPAAGAAL
SEGXX
PRD cceeeeeeccccchhhhhhhccccccccccccccccchhhhhhhccccccccchhhhhc

SEQ LH
SEG ..
PRD cc

(No Prosite data available for DKFZphfbr2_22k8.3)

(No Pfam data available for DKFZphfbr2_22k8.3)

DKFZphfbr2_23b10

group: nucleic acid managment

DKFZphfbr2_2b10 encodes a novel 580 amino acid protein with strong similarity to rat RNA helicase HEL117.

HEL117 is a DEAD/H box helicase, which co-localises with a splicing factor and thus seems to be involved in splicing.

The new protein can find application in modulation of splicing.

strong similarity to rat RNA helicase HEL117

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2905 bp

Poly A stretch at pos. 2885, no polyadenylation signal found

```
1  GGGGGCTCCG CTCCGCACCA CCAACCCCGG GCCGCAGTCC TGACGAGCGG
51  GTCAGGGGCTT GTCGGGCGGA AGCCTGGCCT GGAGCCTGGA AGGGGGAGAC
101 GGGCCGAGCGG GGAGCGGGGAG CGGACGCGGC CTCAGTCTCTG CGCGGAATAT
151 TGAAGGATGT TTGTTCCAAG ATCTCTAAAA ATCAAGAGGA ATGCTAATGA
201 TGATGGCAAA AGTTGTGTGG CTAAGATAAT TAAACCAGAC CCAGAAGACC
251 TTCAGTTGGA CAAAAGCAGA GATGTTCCCG TTGATGCTGT AGCTACAGAA
301 GCAGCCACAA TAGACAGGCA CATCAGCGAA TCATGCCCTT TCCCCAGCCC
351 AGGTGGCCAG TTGGCAGAGG TTCATTCACT AAGTCCCGAG CAGGGTGCGA
401 AGGACAGCCA TCCTTCTGAA GAGCCCGTTA AGTCATTTTC CAAAACACAG
451 CGCTGGGCGAG AACGAGGGGA ACCCATCTGT GTTGTCTGTG GTCGTTATGG
501 ACAGTATATC TGTGATAAGA CAGATGAAGA TGTGTGTAGT TTGGAGTGTA
551 AAGCGAAACA TCTTCTACAA GTTAAGGAAA AGGAAGAGAA ATCAAAACTC
601 AGCAATCCAC AGAAGGCTGA TTCTGAGCCA GAGTCTCCAC TGAATGCTTC
651 CTATGTCTAC AAAGAGCACC CCTTTATTTT GAACCTTCAG GAAGACCAGA
701 TTGAAAATCT TAAACAGCAG CTGGGAATTT TAGTTC AAGG GCAAGAAGTC
751 ACCAGGCCCA TTATTGACTT TGAACATTGT AGTCTCCCTG AGGTCTTAAA
801 TCACAACTTG AAGAAATCAG GCTATGAGGT GCCAACTCCC ATTCAAATGC
851 AGATGATTCC TGTGGGACTT CTGGGAAGAG ACATTCTGGC CAGTGCAGAT
901 ACTGGCTCAG GAAAAACAGC TGCTTTTCTT CTTCTCTGTA TCATGCGAGC
951 TTTATTGCGAG AGCAAAACTC CATCTGCGCT CATTTCTACA CCAACCAGAG
1001 AGTTAGCCAT TCAGATAGAG AGACAAGCTA AAGAATTGAT GAGTGGCCTG
1051 CCACGCATGA AAACGTGTCT TCTGTAGGG GGCTTACCCT TACCCCCACA
1101 GCTTTATCGT CTGCAACAAC ATGTTAAGGT TATCATAGCA ACCCCTGGGC
1151 GACTTCTGGA TATAATAAAG CAGAGCTCTG TAGAACTCTG TGGTGTAAG
1201 ATTGTGGTAG TAGATGAAGC TGATACCATG TTAAGATGG GTTTTCAACA
1251 ACAAGTGCTT GACATTTTGG AAAACATTCC TAATGATTGT CAGACCATTT
1301 TGGTTTCAGC CACAATTCCA ACTAGCATAG AACAGCTAGC AAGCCAGCTT
1351 CTGCATAATC CTGTGAGAAT TATCACTGGA GAAAAGAACC TACCTTGTGC
1401 CAATGTACGT CAGATTATTT TGTGGGTAGA AGACCCAGCC AAAAAGAAAA
1451 AATTATTTGA AATTTTAAAT GATAAGAAAC TCTTTAAGCC TCCAGTGTTA
1501 GTATTGTGGG ACTGCAAACT AGGAGCAGAT CTTTGTAGTG AAGCCGTTC
1551 GAAAATCACA GGGCTGAAAA GCATATCTAT ACATTCTGGG AAGTCGCAAA
1601 TAGAAAGGAA AAACATATTG AAGGGATTAC TTGAAGGAGA CTATGAAGTT
1651 GTAGTGAGCA CAGGAGTCTT GGGACGAGGC CTAGACTTGA TCAGTGTCTG
1701 GCTGGTGTGC AATTTTGATA TGCCTTCAAG TATGGATGAG TATGTCCATC
1751 AGGAAAATAC CTACAAGTCT ACTGGAGGA ATCCCCAGCA TTTTCAACAG
1801 GATGTCAGAA TGACCTTGGG CTATGTTGGC AAAGCACAAT GGAAGAAGA
1851 CAACCAATTG AAGGTCAAAC TAGGCCTTAA AAAAAATTGT TCTTCTAAA
1901 TGAACCTTTA TGTAAGACCC AAGCTTCCTT TATGTAAAA TAGGATACTC
1951 ACTAGGCTTT GGGGCTGACA ATGGTTTTTA AATCTTGCTA ATCTTCCCTG
2001 GAATGAAACC AGCATGACTC AAAGAGAAAA AGAGAGTCTA TAATATTTTC
2051 TAATCCCTGA GTTCTTTTCT TTATATATTA AAAAGGATTA TTAGGCTGGG
2101 TGTGGTGGCT CAGCCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGGGAGT
2151 GGATCACCTC AGTTCGAGAC CAGCCTAACC AACATGGAGA AACCCTGTCT
2201 CTACTAAAAA TACAAAATTA GCCAGGCGTG GTGGCGCATG CCTGTAATCC
2251 CAGCTACTCA GGAGGCTACA GCAGGAGAAT TGCTTGAAC TGGGAGGCAG
2301 AGCCAAGATC GCACCACTGC ACTCCAGCCT GGGCAACAAG AGTGAAACTC
2351 TGCTCAAAA TAATATTAAT GATAATAATA ATAATAATA TAGGGATTAC
2401 TTGCATAATT GTTCTTTTAA AATTATTGGC AGTATTGCTG AATGTATTTA
2451 GATTTTTTCA CCAAGTGACA ACAACTGAAT TCATAAAGAT TCATCAACAA
2501 GACCTGATAA AAAAAATGT AAGCATATTA TAGTGGATAC TTCCAAGACT
2551 CTTGGTCTAA CATGTATTAG AAAGCAGAAG GAGCCCAGGC ACAGGGGCTC
2601 CCGCCGGTAA TCCCAAAGCT TTGGGAAGCC AAGGCAGGTG GATCGCTTGA
2651 GCTCAGGAGT TAGAGACCAG CCTGGGCAAC ATGGTGAAAT CCCGTCACCA
```



```

2701 CAAAAAATG CAAAAATTAA CTGGGCGTGG TGGCATGCAC CTGTAGTCCC
2751 AGCTACTCTG GAGGCTGAGG TGAGGGGAAT CACCTGAGCC GGGGGAATCA
2801 CCTGAGCCCA GGGAAAGTTGA GGCTGCTGTG AGCCATGGTC ATGACACTGC
2851 CCTCCAGCCT GGACAACAGA TTGAGACCCT GTCTCAAAAA AAAAAAAAAA
2901 AAAAA

```

BLAST Results

No BLAST result

Medline entries

Medline:

A putative mammalian RNA helicase with an arginine-serine-rich domain

Peptide information for frame 1

ORF from 157 bp to 1896 bp; peptide length: 580
 Category: strong similarity to known protein
 Prosite motifs: ATP_GTP_A (247-255)
 LEUCINE_ZIPPER (298-320)

```

1 MFVPRSLKIK RNANDDGKSC VAKIIKPDPE DLQLDKSRDV PVDVATEAA
51 TIDRHISESC PFPSPGGQLA EVHSVSPEQG AKDSHPSEEP VKSFSKTQRW
101 AEPGEPICVV CGRYGEYICD KTDEDVCSLE CKAKHLLQVK EKEEKSLSN
151 POKADSEPES PLNASYVYKE HPFILNLQED QIENLKQQLG ILVQGGQEVTR
201 PIIDFEHCSL PEVLNHNLLK SGYEVPTPIQ MQMIPVGLLG RDILASADTG
251 SGKTAFFLLP VIMRALFESK TPSALILTPT RELAIQIERQ AKELMSGSLPR
301 MKTVLLVGGP PLPPQLYRLQ QHVKVIIATP GRLLDIKQS SVELCGVKIV
351 VVDEADTMLK MGFQQQVLDI LENIPNDCQT ILVSATIPTS IEQLASQLLH
401 NPVRIITGEK NLPCANVRQI ILWVEDPAKK KKLFEILNDK KLFKPPVLVF
451 VDCKLGADLL SEAVQKITGL KSISIHSEKS QIERKNILKG LLEGDYEVVV
501 STGVLGRGLD LISVRLVNF DMPSSMDEYV HQENTYKSTW RNPQHQQQDV
551 RMTLGYVGKA QWEEDNQLKV KLGLKNCSS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23b10, frame 1

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 615, P = 1.6e-60

TREMBL:AB018344_1 gene: "KIAA0801"; product: "KIAA0801 protein"; Homo sapiens mRNA for KIAA0801 protein, complete cds., N = 1, Score = 615, P = 2.8e-59

TREMBL:CE01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1., N = 2, Score = 365, P = 1.9e-58

TREMBL:AF083255_1 product: "RNA helicase-related protein"; Homo sapiens RNA helicase-related protein mRNA, complete cds., N = 2, Score = 556, P = 1.5e-57

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces pombe), N = 1, Score = 591, P = 1.6e-57

>PIR:A57514 RNA helicase HEL117 - rat
 Length = 1,032

HSPs:

Score = 615 (92.3 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
 Identities = 140/394 (35%), Positives = 236/394 (59%)

```

Query: 144 EKSLSNPQKADSEPESPLNASYVYKEHPFILNLQEDQIENLKQQL-GILVQGGQEVTRPI 202
      ++ KL P P ++ Y E P + + + + + + + + GI V+G+ +PI
Sbjct: 313 KQRKLLPEVDHGKIEYEPFRKNF-YVEVEPELAKMSQEEVNVFLEMEGITVRGKGCPKPI 371
Query: 203 IDFEHCSLPEVLNHNLLKSGYEVPTPIQMIPVGLLGRDILASADTGSGKTAFFLLPV- 261

```



```

      + C + + ++LKK GYE PTPIQ Q IP + GRD++ A TGSGKT AFLLP+
Sbjct: 372 KSWVQCGISMKILNSLKKHGYEKPITPIQTQAIPAIMSGRDLIGIAKTGSGKTIAFLLPF 431
Query: 262 --IM--RALFESKTPSALILTPTRELAIQERQAKELMSGLPRMKTVLLVGGGLPLPPQLY 317
      IM R+L E + P A+I+TPTRELA+QI ++ K+ L ++ V + GG + Q+
Sbjct: 432 RHIMDQRSLEEGERPIAVIMPTRELALQITKECKKFSKTLG-LRVVCVYGGTGISEQIA 490
Query: 318 RLQQHVKVIIATPGRLLDIIKQSS---VELCGVKIVVVDEADTMLKMGFQQQVLDILENI 374
      L++ ++I+ TPGR++D++ +S L V VV+DEAD M MGF+ QV+ I++N+
Sbjct: 491 ELKRGAEIIVCTPGRMIDMLAANSGRVTNLRRTYVVVLEADRMFDMGFEPQVMRIVDNV 550
Query: 375 PNDCQITLVSATIPTSIQQLASQLLHNPVRIITGEKNLPCANVRQIILWVEDPAKKKKLF 434
      D QT++ SAT P ++E LA ++L P+ + G +++ C++V Q ++ +E+ K KL
Sbjct: 551 RPDQRTVMFSATFFRAMEALARRILSKPIEVQVGGRSVVCSDVEQQVIVIEEEKKFLKLL 610
Query: 435 EILNDKKLKFPPVLVFDCKLGADLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEG 494
      E+L + V++FVD + AD L + + + + +S+H Q +R +I+ G
Sbjct: 611 ELLGHYQE-SGSVII FVDKQEHADGLLKDLMRAS-YPCMSLHGGIDQYDRDSIINDFKNG 668
Query: 495 DYEUVVSTGVLGRGLDLISVRLVVNFDMFSSMDEYVHQ 532
      +++V+T V RGLD+ + LVVN+ P+ ++YVH+
Sbjct: 669 TCKLLVATSVAARGLDVKHLILVVNYSCPNNHYEDYVHR 706

Score = 37 (5.6 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
Identities = 13/36 (36%), Positives = 17/36 (47%)

Query: 132 KAKHLLQVKEKEE---KSKLSNPQKADSEPEPLNA 164
      KA++ + KEK E SK K D E E +A
Sbjct: 113 KAENRSRSKEKAEGDSSKEKKDKDDKEDEKEKDA 148

```

Pedant information for DKFZphfbr2_23b10, frame 1

Report for DKFZphfbr2_23b10.1

```

[LENGTH]      580
[MW]           64572.24
[pI]           6.13
[HOMOL]        TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1. 8e-61

[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]       04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 5e-53
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-49
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-49
[FUNCAT]       j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 2e-46
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 3e-43
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 4e-39
[FUNCAT]       1 genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 3e-35
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 6e-34
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 3e-32
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 8e-30
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 5e-23
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 1e-16
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 5e-11
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 1e-06
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 1e-06
[BLOCKS]       BL00115B Eukaryotic RNA polymerase II heptapeptide repeat proteins
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 6e-53
[PIRKW]        RNA binding 9e-52
[PIRKW]        DEAD box 2e-43
[PIRKW]        transmembrane protein 1e-21
[PIRKW]        DNA binding 5e-48
[PIRKW]        ATP 4e-57
[PIRKW]        purine nucleotide binding 2e-43
[PIRKW]        P-loop 4e-57
[PIRKW]        hydrolase 6e-42
[PIRKW]        protein biosynthesis 2e-43
[PIRKW]        ATP binding 2e-50
[SUPFAM]       WW repeat homology 1e-49
[SUPFAM]       translation initiation factor eIF-4A 2e-43
[SUPFAM]       DEAD/H box helicase homology 4e-57
[SUPFAM]       recQ helicase homology 8e-06

```



```

[SUPFAM]      unassigned DEAD/H box helicases 4e-57
[SUPFAM]      ATP-dependent RNA helicase DBP1 2e-53
[SUPFAM]      ATP-dependent RNA helicase DHH1 6e-40
[SUPFAM]      tobacco ATP-dependent RNA helicase DB10 1e-49
[SUPFAM]      Bloom's syndrome helicase 8e-06
[PROSITE]     ATP_GTP_A 1
[PROSITE]     LEUCINE_ZIPPER 1
[PROSITE]     MYRISTYL 6
[PROSITE]     CK2_PHOSPHO_SITE 8
[PROSITE]     TYR_PHOSPHO_SITE 1
[PROSITE]     PKC_PHOSPHO_SITE 7
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        Helicases conserved C-terminal domain
[PFAM]        DEAD and DEAH box helicases
[KW]          Alpha Beta
[KW]          LOW_COMPLEXITY 3.10 %

```

```

SEQ MFVPRSLKIKRNANDDGKSCVAKIIPDPEDLQLDKSRDVPVDAVATEAATIDRHISESC
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ PFPSPGGQLAEVHSVSPEQGAKDSHPSEEPVKSFSTQRWAEPEGEPICVVCGRYGEYICD
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ KTDEDVCSLECKAKHLLQVKEKEEKSLSNPQKADSEPEPLNASYVYKEHPFILNLQED
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ QIENLKQQLGILVQGGQEVTRPIIDFEHCSLPEVLNHNKSGYEVPTPIQMOMIPVGLLG
SEG .....
PRD hhhhhhhhhheeecccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ RDILASADTSGSKTAAFLLPVIMRALFESKTPSALILPTRELAIQIERQAKELMSGGLPR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ MKTVLLVGGPLPLPPQLYRLQOHVKVVIATPGRLLDIIKQSSVELCGVKIVVDEADTMLK
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ MGFAQQVLDILENIPNDQCITLVSATIPSTIEQLASQLLHNPVRIITGEKNLPCANVRQI
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ ILWVEDPAKKKKLFEILNDKKLFPVVLVFDCKLGADLLSEAVQKITGLKSIHSEKS
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ QIERKNILKGLLEGDYEVVSTGVLGRGLDLISVRLVNFDMPSMDEYVHQENTYKSTW
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ RNPQHQQDVVRMTLGVVGKAQWEEDNQLKVKLGLKKNCS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosite for DKFzphfbr2_23b10.1

PS00001	163->167	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	524->528	CK2_PHOSPHO_SITE	PDOC00006
PS00007	489->497	TYR_PHOSPHO_SITE	PDOC00007
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	80->86	MYRISTYL	PDOC00008

PS00008	195->201	MYRISTYL	PDOC00008
PS00008	250->256	MYRISTYL	PDOC00008
PS00008	490->496	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008
PS00017	247->255	ATP_GTP_A	PDOC00017
PS00029	298->320	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_23b10.1

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAAF		
Query	209	SLPEVLNHNLLKSGYEVPTPIQMqMIPVGLLGRDILASADTSGSKTAAAF	257
HMM	LIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMNgIR		
Query	258	LLPVIMRALFES--KTPS---ALILTPTRELAIQIERQAKELMSGGLPRMK	302
HMM	ImcIYGGtnMRdQMRmLeRGpPHIVATPGRLIDHIERgtldLDRIeMLV		
Query	303	TVLLVGGGLPLPPQLYRLQQHV-KVIIATPGRLLDIIKQSSVELCGVKIVV	351
HMM	MDEADRLMDMGFIDQIRrIMrqIPmpwNRQTMFSATMPdeiQELARrFM		
Query	352	VDEADTMLKMGFQQQVLDILENIP--NDCQTILVSATIPTSIEQLASQLL	399
HMM	RNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdCLcrLlie*		
Query	400	HNPVRIITGEKNLPCA-NVRQIILWVE-DPAKKKKLFEILN	438
HMM_NAME	Helicases conserved C-terminal domain		
HMM	*EileeWLknl.GIrvmyIHGdMpQeERdeIMddFNnGEynVLicTDVgg		
Query	458	DLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEGDYEVVSTGVLG	506
HMM	RGIDIPdVNVHVINYDMPWNPEqYIQRIGRTgRIG*		
Query	507	RGLDLISVRLVVNFDMPPSSMDEYVH-QENTYKST	539

DKFZphfbr2_23b21

group: signal transduction

DKFZphfbr2_23b21.1 encodes a novel 193 amino acid protein which is nearly identical to bovine neurocalcin.

Neurocalcin is a Ca(2+)-binding protein with three putative Ca(2+)-binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.

The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

nearly identical to bovine neurocalcin

complete cds complete cDNA
EST hits

Sequenced by AGOWA

Locus: /map="574.6 cR from top of Chr8 linkage group"

Insert length: 3300 bp
Poly A stretch at pos. 3279, polyadenylation signal at pos. 3249

```
1 GGGGAGAATC TGGTGGATGC TGGACCTTGC TGCTGCTGCT ACTGCTGTTT
51 CCAGGGGGCTG CAGAGCATGG ACTGTTAAAT CTTGCACTTC TTCTGAGTGA
101 GCTGAATTCT TGCCGCCAGG ATGGGGAAAC AGAACAGCAA GCTGCGCCCG
151 GAGGTCATGC AGGACTTGCT GGAAAGCACA GACTTTACAG AGCATGAGAT
201 CCAGGAATGG TATAAAGGCT TCTTGAGAGA CTGCCCCAGT GGACATTTGT
251 CAATGGAAGA GTTTAAGAAA ATATATGGGA ACTTTTCCCT TTATGGGGAT
301 GCTTCCAAAT TTGCAGAGCA TGTCTTCCGC ACCTTCGATG CAAATGGAGA
351 TGGGACAATA GACTTTAGAG AATTCATCAT CGCCTTGAGT GTAACTTCGA
401 GGGGGGAAGCT GGAGCAGAAG CTGAAATGGG CCTTCAGCAT GTACGACCTG
451 GACGGAAATG GCTATATCAG CAAGGCAGAG ATGCTAGTGA TCGTGCAGGC
501 AATCTATAAG ATGGTTTCCT CTGTAATGAA AATGCCTGAA GATGAGTCAA
551 CCCCAGAGAA AAGAACAGAA AAGATCTTCC GCCAGATGGA CACCAATAGA
601 GACGGAAATG TCTCCCTGGA AGAGTTTCATC CGAGGAGCCA AAAGCGACCC
651 GTCCATTGTG CGCCTCCTGC AGTGCGACCC GAGCAGTGCC GGCCAGTTCT
701 GAGCCCTGCG CCCACCAATC GAATTGTAGA GCTGCTTGTG TTCCCTTTTG
751 ATTCTTCTTT TTAACAATTT TTTTTTTTTT TTGCCAACA ATATCAATGG
801 TGATGCCGCT CCCTGTGCGG TCTGATGCGC CTTCCTCCGT GACGCTTCA
851 GCCTCTTTTG TCGTGGATGC TTCGTGGGAA TGCCAGAGC CCCAGTGTG
901 TTGTGGAGAG CATGGACAGA CTCGTGGTG TTCATTGTTT GATGATTTT
951 AATCGTTACT ATTATTCTT TTTATTCTAA TGTCTCTGTT CTAACACGTA
1001 AGACTCGGGG GTTGGGGCAA AAGAAGGGAA ACCCATCCAG TCCTGTGATT
1051 CTATTGCAAG CTTCAAGGGG CTTTTGTTTG AAAGACAAA CTCGCCACCT
1101 GGGTCTGTTG TCACACGTGC CGTAGGGGTG ATGGATGGCA CCGGATGCTG
1151 GATTCCCCAA GAACAAGTTA CCCTCTGGGG TGAGGCTATT CCAGCGAGCT
1201 GGGACATTTC CCCATGGGGG CCCACTCCCC TCTCTTCCCC AGCAGGCTGT
1251 AGTTTCTAAG CTGTGAACAT TTCAGATAA ATTAACAGAG GAGAGGAAAA
1301 AGATGGCTCA GCTATTTTTT CACAGGTTTA CACTAGTTGA GCTAATATGC
1351 GTGTCTTTGG AATTAAACA CAAATGGTAA CATATTCCAA AACCAGACCC
1401 ATCTTGTTGC CTATTGTGAT AAAATAAAAA GACGGCTGTA TATAACATAT
1451 TGGGTAATGC AGACCAAATT AAGTGTTTTG CCTTGTTTAA ATGAAATGCA
1501 TGTTTAGTGA GCACTAATAC AATCTTATC CAGAAGACTG TTTTGTAGT
1551 CTTATTGTGA AGTAAGACAA CTATAATGAA TGTCTGTCTT GTTTGGAAGT
1601 CATATCTGTC TTTGCACAAA TGTACCAATC GACAAGTATA TTTTATATAT
1651 TCCATAAAAA TACAAAGTAA CCCTGACTAG GGCCCAACTT TAATTTTGAA
1701 TGCATTTCCA GAGTGGCCAT GCCTAGAGGG CAGATGCAGA GCAGGTGGTA
1751 GTGGGACAGG ACAATTGGAG CACAGGAATG TTAACATGTA TGACAGGGGA
1801 CCAGTAGGGT GGTTCCTCTC TCAGGCCAG CAGCCCATTG ACAGCATTAG
1851 ACTGGCGGCA TGGTGCTTTT CTGAGCAGAT CAATACTCTG CAGACTCGAA
1901 AAAACATCAC ATACATTCTT GGAACCTCCC AGTGGTTTAA TCTATGTGCA
1951 TGGTTAGGGA GCCAGGCCCTG GAATATTCAG TTTCCCTGCC CCTGTTAAAG
2001 AATCAGAGGT TGGGCAGTCA TCAAAATCAT CATAAAGACA TGGGCAAGTG
2051 TGCTGTGGT TTCCAAGGCC CCCCTATGGA GAATCCAAAA GTATTTTCCA
2101 TTGCCGTGCT CTTTGAATGC AGACTTCTAT TTCCAGAAGT GACAGCACAA
2151 GTCTGAGTTG CTGTTGGTTC TGGTGACCTC AGACACACTA ATTTGAATTG
2201 AAAGCTAAGA GTAAAAATTT GCTGGTTACA GGCGAGTCAT ACTCTTGCAA
2251 GTAGTTAGCA AAGGGAGGCC CAAATTCTCA AGGTTGTTGA TGGGGAACCT
2301 GCCACTAAGA GAAGGCAGAG AGGTCCCTAG TGGGTATATT TGCTGCCAAG
2351 CCCTTGCCA AAGAAGAGGA ACCACAGAAA GAGAGACATC ATGACCAGGA
2401 GAAAAATGTG ACTAGACATG CTAACCTCCA GGTTTTATA TATGACTTGA
2451 GTCTGCTGTA ATTGGCAGCA GAAATCCAAA TTTGTATGGT AGACCAAAAA
2501 GAACCAATC CATAGGGTGA AATTTTGAGA CCTAGACTCT GTAAAAATAA
```



```

2551 TCCTAGTCTT CCTCCAGGGG TCAGTTCCTC ACAGTGGTTC TGTACCAAAA
2601 CTTGCCAAAT TCCTCCATGG CCAAGTGTTA AAATCTGTGT TTGGAATAA
2651 GCGAATTAAC CTAAGACACA GAAGGCAGAC TGGGTGAGGA GACCTAGCAT
2701 GCCCTATTGG CAGTGCTCAG GAGCTGCATC CCACCTTTCC CTGCTCTGAA
2751 TCGAAGTCCT AGTTCCTTCC TTTGATTCTC CTTTGGTAGG TGAATCAGT
2801 TAATGTTTGG AGAAACCTGC CTGGGCTCTG CCCTTAGTCA TGACATCTCG
2851 CTGAGCCAGA CCCACTCTGT TCCTTGGAAC CTAGAGCTGG AGTGAGGAGT
2901 AGAGGTCTCC GGCTATTCCA GAAAGAAAAG TGAGCCACAT GCAGGCTGAT
2951 GAATGCCGAC ACTTCCAGAA TGTATAGAAA TAGTCCCTGT CCTGGCCTGC
3001 CACTGACCTT GTCTGTATTT TCTCGGAGGT TGTTTTTCTC CTTCTCCTTC
3051 CCAGGAAGGT CTTTGTATGT CGAATCCAGT GCACTCAAGT TTGGCCAAGG
3101 GACTCCACAG CACCCAGAGG ACTGCATGCC TCAAGGTTTA TGTCACCTCT
3151 CTGCTGGGCT GTTCATTGTC ATTGCTGTGT TCAGGGACCT TTGGAATAA
3201 AACCTGTCTT GTCCCAAATA AAACCAGCCT GTGATGTTCA AGGGACTGGA
3251 ATAAAGTGGC TTACGACCTG AAGGATTCTA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HS431350 from database EMBL:
human STS WI-15914.
Score = 1308, P = 3.1e-53, identities = 276/285

Entry HSG19929 from database EMBL:
human STS A002C26.
Score = 926, P = 1.5e-35, identities = 186/187

Entry AF052142 from database EMBL:
Homo sapiens clone 24665 mRNA sequence.
Score = 7378, P = 0.0e+00, identities = 1482/1487
3' UTR

Medline entries

93247712:
Neurocalcin family: a novel calcium-binding protein abundant in bovine central nervous system.

94045365:
Distinct regional localization of neurocalcin, a Ca(2+)-binding protein, in the bovine adrenal gland.

96407688:
Crystallization and preliminary X-ray crystallographic studies of recombinant bovine neurocalcin delta.

96066284:
Distribution pattern of three neural calcium-binding proteins (NCS-1, VILIP and recoverin) in chicken, bovine and rat retina.

Peptide information for frame 1

ORF from 121 bp to 699 bp; peptide length: 193
Category: strong similarity to known protein
Prosites motifs: EF_HAND (73-86)
EF_HAND (109-122)
EF_HAND (157-170)

```

1 MGKQNSKLRP EVMQDLLEST DFEHEIQEW YKGFLRDCPS GHLSMEEFKK
51 IYGNFFPYGD ASKFAEHVFR TFDANGDGTI DFREFIIALS VTSRGKLEQK
101 LKWAFFMYDL DNGYISKAE MLVIVQAIYK MVSSVMKMPE DESTPEK RTE
151 KIFRQMDTNR DGKLSLEEFI RGAKSDPSIV RLLQCDPSSA GQF

```

BLASTP hits

Entry JH0616 from database PIR:
neurocalcin (clone pCalN) - bovine

Score = 1001, P = 5.2e-101, identities = 192/193, positives = 192/193

Entry GGU91630_1 from database TREMBL:
product: "neurocalcin"; Gallus gallus neurocalcin mRNA, complete cds.
Score = 998, P = 1.1e-100, identities = 191/193, positives = 192/193

Entry NECD BOVIN from database SWISSPROT:
NEUROCALCIN DELTA.
Score = 996, P = 1.8e-100, identities = 191/192, positives = 191/192

Entry S47565 from database PIR:
BDR-1 protein - human
Score = 934, P = 6.6e-94, identities = 174/193, positives = 187/193

Entry I50676 from database PIR:
gene Rem-1 protein - chicken >TREMBL:GGREM1_1 gene: "Rem-1"; G.gallus
rem-1 mRNA
Score = 933, P = 8.4e-94, identities = 174/193, positives = 186/193

Alert BLASTP hits for DKFZphfbr2_23b21, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23b21, frame 1

Report for DKFZphfbr2_23b21.1

```
[LENGTH]      193
[MW]           22215.30
[pI]           5.35
[HOMOL]        PIR:JH0616 neurocalcin (clone pCalN) - bovine 1e-109
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YDR373w] 3e-54
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[FUNCAT]       [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       03.01 cell growth [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       13.04 homeostasis of other ions [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       08.19 cellular import [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR109c]
0.001
[FUNCAT]       10.02.99 other morphogenetic activities [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       30.05 organization of centrosome [S. cerevisiae, YBR109c] 0.001
[BLOCKS]       BL00018
[SCOP]         dlrec_ 1.34.1.5.18 Recoverin (bovine (Bos taurus) 8e-55
[SCOP]         dljsa_ 1.34.1.5.17 Recoverin (human (Homo sapiens) 5e-58
[SCOP]         dlrcob_ 1.34.1.5.16 Calcineurin regulatory subunit (B-chain 1e-06
[SCOP]         d2mysc_ 1.34.1.5.15 Myosin Regulatory Chain (chicken (Gallu 2e-29
[SCOP]         dlscmc_ 1.34.1.5.14 Myosin Regulatory Chain (bay scallo 5e-33
[SCOP]         d2mysb_ 1.34.1.5.13 Myosin Essential Chain (chicken (Gallu 4e-26
[SCOP]         dlscmb_ 1.34.1.5.12 Myosin Essential Chain (bay scallo 6e-27
[SCOP]         dlclm_ 1.34.1.5.11 Calmodulin (Paramecium tetraurelia 1e-15
[SCOP]         d4cln_ 1.34.1.5.10 Calmodulin (Drosophila melanogaster 2e-16
[SCOP]         dlcf_ 1.34.1.5.9 Calmodulin (African frog (Xenopus laevis) 2e-16
[SCOP]         dlahr_ 1.34.1.5.8 Calmodulin (chicken gallus gallus 4e-16
[SCOP]         d3cln_ 1.34.1.5.7 Calmodulin (rat (Rattus rattus) 2e-16
[SCOP]         dltrcb_ 1.34.1.5.6 Calmodulin (bovine (Bos taurus) 8e-08
[SCOP]         dlcll_ 1.34.1.5.5 Calmodulin (human (Homo sapiens) 2e-16
[SCOP]         dlrtpl_ 1.34.1.4.5 Parvalbumin (rat (Rattus rattus) 8e-06
[SCOP]         d5tnc_ 1.34.1.5.2 Troponin C (turkey (Meleagris gallopavo) 3e-13
[SCOP]         dlpvaa_ 1.34.1.4.3 Parvalbumin (pike (Esox lucius) 6e-06
[SCOP]         dltnp_ 1.34.1.5.1 Troponin C (chicken (Gallus gallus) 9e-11
[EC]           2.7.1.107 Diacylglycerol kinase 2e-08
[PIRKW]        blocked amino end 1e-100
[PIRKW]        phosphotransferase 2e-08
[PIRKW]        duplication 4e-17
[PIRKW]        tandem repeat 7e-06
[PIRKW]        heterodimer 4e-17
[PIRKW]        heart 6e-09
[PIRKW]        zinc 2e-08
[PIRKW]        serine/threonine-specific protein kinase 1e-06
[PIRKW]        muscle contraction 1e-08
[PIRKW]        acetylated amino end 4e-09
[PIRKW]        ATP 2e-08
[PIRKW]        skeletal muscle 6e-09
```



```

[PIRKW]      signal transduction 1e-91
[PIRKW]      protein kinase 2e-08
[PIRKW]      calcium binding 1e-100
[PIRKW]      alternative splicing 2e-13
[PIRKW]      methylated amino acid 1e-09
[PIRKW]      thin filaments 1e-08
[PIRKW]      lipoprotein 1e-101
[PIRKW]      cardiac muscle 6e-09
[PIRKW]      muscle 6e-09
[PIRKW]      myristylation 1e-100
[PIRKW]      EF hand 1e-101
[PIRKW]      retina 2e-51
[SUPFAM]     calcium-dependent protein kinase 2e-08
[SUPFAM]     unassigned calmodulin-related proteins 8e-41
[SUPFAM]     spec-related protein LpS1 7e-06
[SUPFAM]     calmodulin repeat homology 1e-101
[SUPFAM]     human diacylglycerol kinase 2e-08
[SUPFAM]     protein kinase C zinc-binding repeat homology 2e-08
[SUPFAM]     protein kinase homology 2e-08
[SUPFAM]     calmodulin 1e-101
[PROSITE]    EF_HAND 3
[PROSITE]    CK2_PHOSPHO_SITE 7
[PROSITE]    PKC_PHOSPHO_SITE 3
[PFAM]       EF hand
[KW]         All_Alpha
[KW]         3D

```

```

SEQ      MGKQNSKLRPVMDLLESTDFTEHEIQEWYKGLRDCPSGHLSEEFKKIYGNFFPYGD
1rec-    .....HHHHHHHHHTTTCCCHHHHHHHHHHHHHHTTTTEEEHHHHHHHHHHHTTTTC

SEQ      ASKFAEHVFTFDANGDGTIDFREIIALSVTSRGKLEQKLKWFMSMYDLGNGYISKAE
1rec-    HHHHHHHHHHHH-----CEEHHHHHHHHHHHHHCCCGGHHHHHHHHHTTTTCCCEEHHH

SEQ      MLVIVQAIYKMVSSVMKMPEDSTPEKRTEKIFRQMDTNRDGKLSLEEFIRGAKSDPSIV
1rec-    HHHHHHHHHHCCCTTGGGCTTTTCHHHHHHHHHHHHCCCTTTTECHHHHHHHHHHCHHHH

SEQ      RLLQCDPSSAGQF
1rec-    HHHCCCH.....

```

Prosite for DKFZphfbr2_23b21.1

PS00005	92->95	PKC_PHOSPHO_SITE	PDOC00005
PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	44->48	CK2_PHOSPHO_SITE	PDOC00006
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	158->162	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00018	73->86	EF_HAND	PDOC00018
PS00018	109->122	EF_HAND	PDOC00018
PS00018	157->170	EF_HAND	PDOC00018

Pfam for DKFZphfbr2_23b21.1

```

HMM_NAME      EF hand
HMM            *MFrmMDkDGDGyIDFEEmEMMkem*
               +FR +D +GOG+IDF EF+ +++
Query          68  VFRTFDANGDGTIDFREIIALSVT          92

30.75   100   128       1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
Query          *EIqEMFrmMDkDGDGyIDFEEmEMMkem*
               ++++F+M+D DG+GYI++ E+++++++
dkfzphfbr2    100   KLKWFMSMYDLGNGYISKAEMLVIVQAI    128

Query          176       1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
HMM            *EIqEMFrmMDkDGDGyIDFEEmEMMkem*
               +++FR MD+++DG+++ EEf++ K+
Query          148   RTEKIFRQMDTNRDGKLSLEEFIRGAKSD    176

```


DKFZphfbr2_23f2

group: brain derived

DKFZphfbr2_23f2 encodes a novel 182 amino acid protein with weak similarity to *S. pombe* Vps29p.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Vps29p

complete cDNA, complete cds, EST hits
S. cerevisiae and *S. pombe* Vps29p are involved in vacuolar protein sorting
 part of the cDNA is encoded by HSAC2350, splice pattern 4 exons

Sequenced by AGOWA

Locus: /map="12q24"

Insert length: 1016 bp

Poly A stretch at pos. 996, polyadenylation signal at pos. 974

```

1 GAATGGGGAG GAGCCAGAGG AAGAGGGCGG CGACGGTGGT GGTGACTGAG
51 CGGAGCCCGG TGACAGGATG TTGGTGTGGT TATTAGGAGA TCTGCACATC
101 CCACACCGGT GCAACAGTTT GCCAGCTAAA TTCAAAAAC TCCTGGTGCC
151 AGGAAAAATT CAGCACATTC TCTGCACAGG AAACCTTTGC ACCAAAGAGA
201 GTTATGACTA CCTCAAGACT CTGGCTGGTG ATGTTTCATAT TGTGAGAGGA
251 GACTTCGATG AGAATCTGAA TTATCCAGAA CAGAAAGTTG TGAAGTTGG
301 ACAGTTCAAA ATTGGTCTGA TCCATGGACA TCAAGTTATT CCATGGGGAG
351 ATATGGCCAG CTTAGCCCTG TTGCAGAGGC AATTTGATGT GGACATTCTT
401 ATCTCGGGAC ACACACACAA ATCTGAAGCA TTTGAGCATG AAAATAAATT
451 CTACATTAAT CCAGGTTCTG CCACTGGGGC ATATAATGCC TTGAAACAA
501 ACATTATTC ATCATTGTG TTGATGGATA TCCAGGCTTC TACAGTGGTC
551 ACCTATGTGT ATCAGCTAAT TGGAGATGAT GTGAAAGTAG AACGAATCGA
601 ATACAAAAA CCTTAAAGCC AGGCCTGTCT TGATGATTTT TGGTTTTTTT
651 TCATGTCTCT GTTGAATCA AGTAATTAAA CATTAAAGAG CCACAAAAAT
701 GTATCACTTT TATAATATTT TGCAGTAAAA TATAATACCA TCTTCTCTGT
751 TAATACATAA TTGCTCCAAG CTTCCTGTAA ACTATAAGAA TATATTAGT
801 TTACAGTATA TGGATTCTAT GAAAAAATGT CCACAACACA GTAATTGGTC
851 ACTTGTTAAG AAAAAATTTAT CCTTGTAAGT ATCTTCAAAG TTGATATTTG
901 GAACCTTTAT CAAAAGTAG TGCATGTGGA GAAAGAATCT AGACTTTCTT
951 GTATACATTT TTCTCTTCTC CAGTAATAAA CAATTACCTT TCATTGAAAA
1001 AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HSAC2350 from database EMBLNEW:
 Homo sapiens 12q24 PAC P424M6 Length = 167,217

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 68 bp to 613 bp; peptide length: 182
 Category: similarity to known protein
 Prosite motifs: RGD (60-63)

```

1 MLVLVLGLDLH IPHRCNSLPA KFKLLVPGK IQHILCTGNL CTKESYDYLK
51 TLAGDVHIVR GDFDENLNYP EQKVVTVGQF KIGLIHQHV IPWGDMA SLA
101 LLQRQFDVDI LISGHTHKSE AFEHENKFYI NPGSATGAYN ALETNIIPSF

```


151 VLMDIQASTV VTYVYQLIGD DVKVERIEYK KP

BLASTP hits

Entry CEZK1128_6 from database TREMBL:
 "ZK1128.1"; *Caenorhabditis elegans* cosmid ZK1128
 Length = 523
 Score = 400 (140.8 bits), Expect = 2.3e-37, P = 2.3e-37
 Identities = 81/150 (54%), Positives = 106/150 (70%)

Entry S46793 from database PIR:
 hypothetical protein YHR012c - yeast (*Saccharomyces cerevisiae*)
 Length = 282
 Score = 180 (63.4 bits), Expect = 3.7e-37, Sum P(3) = 3.7e-37
 Identities = 35/71 (49%), Positives = 44/71 (61%)

Entry AB011824_1 from database TREMBL:
 "Vps29"; *Schizosaccharomyces pombe* mRNA for Vps29,
 partial cds. *Schizosaccharomyces pombe* (fission yeast)
 Length = 176
 Score = 189 (66.5 bits), Expect = 2.7e-27, Sum P(2) = 2.7e-27
 Identities = 33/72 (45%), Positives = 50/72 (69%)

Alert BLASTP hits for DKFZphfbr2_23f2, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23f2, frame 2

Report for DKFZphfbr2_23f2.2

[LENGTH] 182
 [MW] 20445.84
 [pI] 6.29
 [HOMOL] TREMBL:CEZK1128_6 gene: "ZK1128.8"; *Caenorhabditis elegans* cosmid ZK1128 2e-51

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YHR012w] 1e-27
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YHR012w] 1e-27
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YHR012w] 1e-27
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YHR012w] 1e-27
 [FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YHR012w] 1e-27
 [FUNCAT] r general function prediction [M. jannaschii, MJ0623] 1e-16

[BLOCKS] BL01269D
 [BLOCKS] BL01269A
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 4
 [PROSITE] PKC_PHOSPHO_SITE 1
 [KW] Alpha_Beta

SEQ MLVLVLGDLHIPHCNSLPKFKLLVPGKIQHILCTGNLCTKESYDYLKTLAGDVHIVR
 PRD cccceccccccccccccchhhhhhhhhccceeeccccccchhhhhhhhhhhccceee

SEQ GDFDENLNYPEQKVVTGQFKIGLIHGQVIPWGDMSALLQRFQDVIDILISGHTHKSE
 PRD cccccccccccccccccccccccccccccchhhhhhhhhhhccceeecccccccc

SEQ AFEHENKFYINPGSATGAYNALETNIIPSFLMDIQASTVVTYVYQLIGDDVKVERIEYK
 PRD ccc

SEQ KP
 PRD cc

Prosite for DKFZphfbr2_23f2.2

PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	83->89	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	137->143	MYRISTYL	PDOC00008
PS00016	60->63	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_23f2.2)

DKFZphfbr2_23124

group: intracellular transport and trafficking

DKFZphfbr2 23124.2 encodes a novel 348 amino acid protein with similarity to human glycoprotein gp36b and canine VIP36 glycoprotein.

The vesicular protein VIP36 (36 kDa vesicular integral membrane protein) shows homology to leguminous plant lectins. The protein is localized to the Golgi apparatus, endosomal and vesicular structures and the plasma membrane. VIP36 binds to sugar residues of glycosphingolipids and/or glycosylphosphatidyl-inositol anchors and might provide a link between the extracellular/luminal face of glycolipid rafts and the cytoplasmic protein segregation machinery. Gp36 is located within the endoplasmic reticulum. For the novel protein, a lectin character is predicted. Due to the intracellular localisation of the homolog proteins, it should be involved in intracellular transport and trafficking.

The new protein can find application in modulating/blocking intracellular transport and trafficking.

strong similarity to human GP36b glycoprotein

complete cDNA, complete cds, EST hits
potential start at Bp 29 matches kozak consensus ANNatgG
similarity to lectins,

Sequenced by AGOWA

Locus: /map="2"

Insert length: 2416 bp

Poly A stretch at pos. 2394, no polyadenylation signal found

```

1  GGGGGATGAA GGGTCGTTGG TGGGAAAGAT GCGCGCGACT CTGGGACCCC
51  TTGGGTCGTG GCAGCAGTGG CCGCGATGTT TGTCGGGCTCG GGATGGGTCC
101 AGGATGTTAC TCCTTCTTCT TTTGTTGGGG TCTGGGCAGG GGCCACAGCA
151 AGTCGGGGCG GGTCAAACGT TCGAGTACTT GAAACGGGAG CACTCGCTGT
201 CGAAGCCCTA CCAGGGTGTG GGCACAGGCA GTTCCTCACT GTGGAATCTG
251 ATGGGCAATG CCATGGTGAT GACCCAGTAT ATCCGCCTTA CCCCAGATAT
301 GCAAAGTAA CAGGGTGCCT TGTGGAACCG GGTGCCATGT TTCTGAGAG
351 ACTGGGAGTT GCAGGTGCAC TTCAAATCC ATGGACAAGG AAAGAAGAAT
401 CTGCATGGGG ATGGCTTGGC AATCTGGTAC ACAAGGATC GGATGCAGCC
451 AGGGCCTGTG TTTGGAAACA TGGACAAATT TGTGGGGCTG GGAGTATTTG
501 TAGACACCTA CCCCAATGAG GAGAAGCAGC AAGAGCGGGT ATTCCCTAC
551 ATCTCAGCCA TGGTGAACAA CGGCTCCCTC AGCTATGATC ATGAGCGGGA
601 TGGGCGGCCT ACAGAGCTGG GAGGCTGCAC AGCCATTGTC CGCAATCTTC
651 ATTACGACAC CTTCTGGTG ATTCTGCTACG TCAAGAGGCA TTTGACGATA
701 ATGATGGATA TTGATGGCAA GCATGAGTGG AGGGACTGCA TTGAAGTGCC
751 CGGAGTCCGC CTGCCCCGCG GCTACTACTT CGGCACCTCC TCCATCACTG
801 GGGATCTCTC AGATAATCAT GATGTCAATT CTTGAAGTT GTTTGAAGTG
851 ACAGTGGAGA GAACCCAGAG AGAGGAAAAG TCCTATCGAG ATGTGTTCTT
901 GCCTCAGTGG GACAATATGA AGCTGCCTGA GATGACAGCT CCACTGCCGC
951 CCCTGAGTGG CTTGGCCCTC TTCCTCATCG TCTTTTCTC CTTGGTGT
1001 TCTGTATTTG CCATAGTCAT TGGTATCATA CTCTACAACA AATGCGAGGA
1051 ACAGAGCCGA AAGCGCTTCT ACTGAGCCCT CTTGCTGCCA CCCTTTTGT
1101 GACTGTCAAC CATGAGGTAT GGAAGGAGCG GGCCTGGCC TGAGCATGCA
1151 GCCTGGAGAG TGTCTTGTC TCTAGCAGCT GGTGGGGGAC TATATTCTGT
1201 CACTGGAGTT TTGAATGCAG GGACCCCGCA TTCCATGGT TGTGCATGGG
1251 GACATCTAAC TCTGGTCTGG GAAGCCACCC ACCCCAGGGC AATGCTGCTG
1301 TGATGTGCCT TTCCTGCAG TCCTTCCATG TGGGAGCAGA GGTGTGAAGA
1351 GAATTACGT GGTGTGATG CCAAATCAC GGAACAGAAT TTCATAGCCC
1401 AGGCTGCCGT GTTGTGAC TCAGAAGGCC CTTCTACTTC AGTTTGAAT
1451 CCACAAAGAA TTAATAACTG GTAACACCAC AGGCTTTCTG ACCATCCATT
1501 CGTTGGGTTT TGCATTTGAC CCAACCTCT GCCTACCTGA GGAGCTTTCT
1551 TTGGAACCA GATGGAAC TTCTTCCCTG CCTTACCTTC CTTTCACTCC
1601 ATTCAATTGT CTCTCTGTGT GCAACCTGAG CTGGGAAAGG CATTTGGATG
1651 CCTCTCTGTT GGGCCCTGGG GCTGCAGAAC ACACCTGCCT TTCGCTGGCC
1701 TTCATTAGGT GGCCCTAGGG AGATGGCTTT CTGCTTTGGA TCACTGTTCC
1751 CTAGCATGGG TCTTGGTCT ATTGGCATGT CCATGGCCTT CCCAATCAAG
1801 TCTCTTCAGG CCCTCAGTGA AGTTTGGCTA AAGGTTGGTG TAAAAATCAA
1851 GAGAAGCCTG GAAGACACCA TGGATGCCAT GGATTAGCTG TGCAACTGAC
1901 CAGCTCCAGG TTTGATCAAA CCAAAGCAA CATTGTGCAT GTGGTCTGAC
1951 CATGTGGAGA TGTTTCTGGA CTTGCTAGAG CTTGCTTAGC TGCATGTTT
2001 GTAGTTACGA TTTTGGAAAT CCCTCTTGA GTGCTGAAAG TGAAGGAAG
2051 CTTTCTTCTT ACACCTTGGG CTTGGATATT GCCCAGAGAA GAAATTTGGC
2101 TTTTCTTCTT TAATGGACAA GGGACAGTTG CTGTTCTCAT GTTCCAAGTC
2151 TGAGAGCAAC AGACCTCAT CATCTGTGCC TGGAAAGATT CACTGTCATT
2201 GAGCAGCACA GCCTGAGTGC TGGCCTCTGT CAACCTTAT TCCACTGCCT

```


2251 TATTTGACAA GGGGTTACAT GCTGCTCACC TTACTGCCCT GGGATTAAAT
 2301 CAGTTACAGG CCAGAGTCTC CTTGGAGGGC CTGGAACCTCT GAGTCCTCCT
 2351 ATGAACCTCT GTAGCCTAAA TGAAATTCTT AAAATCACCG ATGGAACCAA
 2401 AAAAAAAAAA AAAAAA

BLAST Results

Entry HS622145 from database EMBL:

human STS WI-6746.

Score = 1079, P = 5.1e-43, identities = 219/223

Entry G42541 from database EMBLNEW:

SHGC-58649 Human Homo sapiens STS genomic, sequence tagged site.

Score = 1091, P = 1.7e-43, identities = 219/220

Medline entries

94265253:

A putative novel class of animal lectins in the secretory pathway
 homologous to leguminous
 lectins.

94208543:

VI36, a novel component of glycolipid rafts and exocytic carrier
 vesicles in epithelial cells.

Peptide information for frame 2

ORF from 29 bp to 1072 bp; peptide length: 348
 Category: strong similarity to known protein

1 MAATLGPLGS WQWRRCLSA RDGSRMLLLL LLLGSGQGPO QVGAGQTFEY
 51 LKREHSLSKP YQGVGTGSSS LWNLMGNAMV MTQYIRLTPD MQSKQCALWN
 101 RVPCFLRDWE LQVHFKINGQ GKKNLHGDGL AIWYTKDRMQ PGPVFGNMDK
 151 FVGLGVFVDT YPNEEKQQR VFPYISAMVN NGSLSYDHER DGRPTLGGC
 201 TAIVRNLYHD TFLVIRYVKR HLTIMMDIDG KHEWRDCIEV PGVRLPRGYY
 251 FGTSSITGDL SDNHDVISLK LFELTVERTP EEEKLHRDVF LPSVDNMKLP
 301 EMTAPLPPLS GLALFLIVFF SLVFSVFAIV IGIILYNKQ EQSRKREFY

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23124, frame 2

PIR:G01447 GP36b glycoprotein - human, N = 1, Score = 1001, P =
 5.9e-101

SWISSPROT:VP36_CANFA VESICULAR INTEGRAL-MEMBRANE PROTEIN VIP36
 PRECURSOR (VIP36)., N = 1, Score = 990, P = 8.6e-100

TREMBL:CET04G9_2 gene: "T04G9.3"; Caenorhabditis elegans cosmid
 T04G9., N = 1, Score = 614, P = 6e-60

PIR:S42626 ER-golgi intermediate compartment protein - human, N = 2,
 Score = 397, P = 1e-42

>PIR:G01447 GP36b glycoprotein - human
 Length = 356

HSPs:

Score = 1001 (150.2 bits), Expect = 5.9e-101, P = 5.9e-101
 Identities = 197/356 (55%), Positives = 256/356 (71%)

Query: 1 MAATLGPLGSWQWRRCLSDRG-----SRMLLLLLLLGSGQGPPQVGAGQTFEYLK 52
 MAA G + W RRCL R G + L LLLLLGS + G + E+LK
 Sbjct: 1 MAAE-GWIWRWGWRRCLG-RPGLLPGPGPTTFLLLLLGSVTA--DITDGNs-EHLK 55

Query: 53 REHSLSKPYQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQKQALWNRVPCFLRDWELQ 112
 REHSL KPYQGVG+ S LW+ G+ M+ +QY+RLTPD +SK+G++WN PCFL+DWE+
 Sbjct: 56 REHSLIKPYQGVGSSSMLWDFQGSTMLOTSQYVRLTPDERSKEGSIWNHQPCLKDWEMH 115

Query: 113 VHFKEHGGQKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFKVGLGVFVDTPNEEKQQERVF 172
 VHFKEHGG QKKNLHGDG+A+WYT+DR+ PGPVFG+ D F GL +F+DTPN+E ERVF
 Sbjct: 116 VHFKEHGGTGGKKNLHGDGLAIWYTRDLVPGPVFGSKDNFHLAIFLDTPNDETT-ERVF 174

Query: 173 PYISAMVNNNGSLSYDHERDGRPTLGGCTAIVRNLYHDTFLVIRYVVRHLTIMMDIDGKH 232
 PYIS MVNNGSLSYDH +DGR TEL GCTA RN +DTFL +RY + LT+M D++ K+
 Sbjct: 175 PYISVMVNNNGSLSYDHSKDRWTELACTADFRNRDHTFLAVRYSRGLTVMTDLEDKN 234

Query: 233 EWRDCIEVPVGRVLPGRYYFGTSSITGDLSDNHDVISLKLFEELTVERTPEEEKLHRDVFLP 292
 EW++CI++ GVRLP GYYFG S+ TGDLSNHD+IS+KLF+L VE TP+EE + P
 Sbjct: 235 EWKNCIDITGVRVLPGRYYFGASAGTGDLSNHDIIISMKLFQLMVEHTPDEESIDWTKIEP 294

Query: 293 SVDNMKLPMTAPLP-----PLSGLALFLIVFSLVFSVFAIVIGIILYNKWEQSRK 345
 SV+ +K P+ P PL+G +FL++ +L+ V V+G +++ K QE++ K
 Sbjct: 295 SVNFKSPKDNVDPTGNFRSGPLTGWRVFLLLLCALLGIVVCAVVGAVVFQKRQERN-K 353

Query: 346 RFY 348
 RFY
 Sbjct: 354 RFY 356

Pedant information for DKFZphfbr2_23124, frame 2

Report for DKFZphfbr2_23124.2

[LENGTH] 348
 [MW] 39711.10
 [pI] 8.55
 [HOMOL] PIR:G01447 GP36b glycoprotein - human 1e-101
 [PIRKW] lectin 2e-37
 [PIRKW] transmembrane protein 2e-37
 [PIRKW] endoplasmic reticulum 2e-37
 [PIRKW] Golgi apparatus 2e-37
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 5
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] Alpha_Beta
 [KW] SIGNAL_PEPTIDE 39
 [KW] LOW_COMPLEXITY 7.76 %

SEQ MAATLGLPGSWQWRCLSDRGSRMLLLLLLLGSGQGPQVAGQTFEYLKREHSLSKP
 SEGxxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhccccccccccccchhhhhhhhhcccc

SEQ YQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQKQALWNRVPCFLRDWELQVHFKEHGG
 SEG
 PRD cccccccccceccccccccceeeccccchhhhhccccccccccccchhhhhhhheeeccc

SEQ GKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFKVGLGVFVDTPNEEKQQERVFYISAMVN
 SEG
 PRD cccccccccceeeccccccccccccccccccccceeeccccccccccccccccceeeccc

SEQ NGSLSYDHERDGRPTLGGCTAIVRNLYHDTFLVIRYVVRHLTIMMDIDGKHEWRDCIEV
 SEG
 PRD cccccccccccccccccccccccccccccccccceeehhhhhhheeecccccccccccccc

SEQ PGVRLPRGYYFGTSSITGDLSDNHDVISLKLFEELTVERTPEEEKLHRDVFLPSVDNMKLP
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccc

SEQ EMTAPLPPLSGLALFLIVFSLVFSVFAIVIGIILYNKWEQSRKRKY
 SEGxxxxxxx.....
 PRD cccccccccchhhcccc

Prosites for DKFZphfbr2_23124.2

PS00001	181->185	ASN_GLYCOSYLATION	PDOC00001
PS00002	35->39	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005

PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00009	120->124	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23124.2)

DKFZphfbr2_23n16

group: signal transduction

DKFZphfbr2_23n16.1 encodes a novel 292 amino acid protein with weak similarity to putative phosphatidylinositol-4-phosphate 5-kinase of *Arabidopsis thaliana*.

The novel proteins contains a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore the new protein should be involved in intracellular signal transduction.

The new protein can find application in modulating/blocking intracellular signal transduction pathways.

similarity to putative phosphatidylinositol-4-phosphate 5-kinase

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2936 bp

Poly A stretch at pos. 2916, polyadenylation signal at pos. 2873

```
1  GGGGGCGCTC  CCGAGAAAGA  GTGAGGGCGC  GACGCGCACC  AACGGTGGAG
51  GGATGTTTCA  GCAGCCCCCTG  AGAAGGAAGA  GGAGGAAGCT  GAGGGCCCCG
101 TGAGGGCGCA  GGACCTGAGG  GAGTCCTACA  TCCAGCTCGT  CCAGGGGTGTG
151 CAGGAGTGGC  AGGATGGTTG  CATGTACCAG  GGGGAGTTTG  GGTGTAACAT
201 GAAGCTTGGA  TATGGCAAAT  TCTCTGGGCC  CACAGGCGAG  TCATACCATG
251 GGCAGTTTTA  CCGGGACCAC  TGCCATGGCC  TGGGTACCTA  CATGTGGCCA
301 GATGGCTCCA  GTTTCACGGG  CACATTTTAC  CTCAGCCACC  GAGAAGGCTA
351 CGGCACCATG  TACATGAAGA  CACGGCTTTT  CCAGACTCAC  TGCCACAACG
401 ACATTGTCAA  CCTTCTCCTG  GACTGTGGGG  CCGACGTGAA  CAAGTGCTCA
451 GATGAGGGTC  TCACGGCACT  CAGCATGTGT  TTCCTCCTCC  ACTACCCCGC
501 CCAGTCCCTT  CAGCCCAATG  TTGCTGAACG  GACCATACCT  GAGCCCCAGG
551 AACCTCCAAA  ATTCCCAAGT  GTTCCAATCC  TTTCATCATC  ATTTATGGAC
601 ACAACCTGG  AGTCTCTGTA  CTATGAGGTG  AACGTGCCTT  CCCAGGGTAG
651 CTATGAGCTG  AGGCCACCGC  CAGCACCCT  GCTCCTGCCA  CGCGTCTCAG
701 GCAGCCACGA  GGGCGGCCAC  TTCCAGGACA  CCGGGCAGTG  TGGGGGGTCC
751 ATAGACCACA  GGAGCAGCTC  TCTGAAGGGG  GACTCCCGT  TGGTGAAGGG
801 CAGCCTTGGC  CATGTGGAAA  GCGGGCTTGA  GGACGTGTTG  GGAGACACAG
851 ACCGGGGCAG  TCTGTGCACT  GCTGAGACGA  AATTTGAGTC  CAACTTGTGT
901 GTGTGCGACT  TCTCCATCGA  GCTCTCGCAG  GCCATGCTGG  AGAGAAGCGC
951 CCAGTCCCAC  AGCTTGCTGA  AGATGGCCTC  GCCCTCACCG  TGCACCAGCA
1001 GCTTCGACAA  AGGGACCATG  CGGAGGATGG  CGCTGTCCAT  GATCGAGTAG
1051 GTCCTGGCAC  CAGCTGGTGG  GGGTGGAGGG  CCACCATCAG  GGCTGAATCC
1101 TATGCTCAGC  AGACCCACGT  CTCTCCCTG  TGCCAGTGGG  AGGCGTTGTG
1151 TCTGGAGATG  TGTGTCTGAA  TGTGTGAGCA  TCCCTGTGTC  GGTGGCTCCA
1201 TGCCATGGCC  AGCCCTGTGG  GGGTGCCACG  GTGACGGGCT  GTTTTCAGTG
1251 CCACCCAGC  CCTGTGGGGG  TGCCACGGTG  ACGGGCTGTT  TTCAGTACCA
1301 CGCCAGCCCT  GCTTTGGCCT  TTGGCACTGG  CCTGAAGTGT  CTCGTGGGA
1351 GCCTCAGCAG  GGGCCACTGT  CAGGGGTCC  ATCCTAGCCA  TAGTGACGCT
1401 GAGTGACACC  TGCCCTGGCA  GCTCTCACAC  CCCTGTGTG  CACCTGTCT
1451 ATACCAAGTG  GTCTCAAAAT  GTGGTCTATG  CACCCCGGG  GGTCCAAGAC
1501 CCTTTCAGGG  AGTCTGTGGG  GTCAAAATGA  TTCTCTGAT  AACCTGAGA
1551 CTCTGTAGC  CTTCTCCTTG  TGTGTATGTT  GGTGGATGGT  ATGAAGACAG
1601 GGCCGTGCAG  ACCACAGCC  CCCAGCGTGC  AGGCGACGAG  TGCCCGGCT
1651 GCTTGGGGGC  ATGGTATTCC  TTCACCACGG  TGTGCACTTG  CGGGGATGCC
1701 TGTCTCACTG  AAGAATGCCT  TTGACTAAGC  AGAAAAGCAA  TGACAAATTG
1751 CATTAATCT  TGCTCTTTC  GTACACACCC  CTCGAATATT  CTGGGTCGGA
1801 AAACATGGGA  AGGACACTGA  TGTGTGCTG  CCACAGACCA  AGGCACACCG
1851 CTTCCCGCA  AGAAGCGCTT  CCCCCAGGGC  CAGAGTAGCA  ACAGAATGCG
1901 GCATCTTCCC  AACCTCCTGC  CCCATTTTGG  ATTGGAAGAA  TGACCACTGG
1951 TATGTGGCTG  TTCATTCTCC  TGAACACAGC  CTGCCACTTT  AAGGAAAACA
2001 TATGACACTA  TTTGTTGCTG  GCGAAATTTA  CATTTTCAAG  TGAATAGCAG
2051 AATTCTGGAC  ACTTGCCACC  ACCACCAAAA  CCTTCATAGC  TTCCCTTAAC
2101 TTTGAGACAT  GGGTGTTCAG  AGGTTTTTCA  CGTGAGATGG  CGTTAGCAGC
2151 GCAGTTTTGT  GATACTGCCT  GAAGACATGC  CGACAGTGCC  CAGATCTCTT
```



```

2201 CTATTGGTGA GCCAGCTTTT CCCACACGGC CAAGTTCTGA TGTGAACCA
2251 TTGCCAGGTG GGTGAAGATC CATTGACAGT GAGAGGTGGG CCCGTGGGCT
2301 TCAGTGCAGC CAGGCGCAGA AGGCTGGTTC ATGAGTGTCC AGCTCCGCCA
2351 GGTAGCTAGC TCACCACCCC CAGCCTGGGT TCATGTAGTT CAAATAGGAA
2401 GACCACGATG ATCAGAAAGG CTGCTCAAAT ACTCCTTCGT CCAGCCGCGT
2451 ACCTGGGGGA GGCTGAATCT CCACTCACTT CCACCAAGGC TGTGCAGAGC
2501 AGATAGGGGA ATCCAGCAAA GGTGGAAAC AGTGCCATCC TTCTCCCCAA
2551 CTGGTTTTGT TTTGTAAAT AACTTTTTGT GACAGTGTTA CTTATTAGTA
2601 ACATGCAGTG GGTTTGTTAT GGTAAACAAG TTGGTGAGCA TTATTGAGAG
2651 GTGAAGCCAG CTGAGCTTCT GGGTTGGGTG GGGACTTGGA GAACTTTTGT
2701 GTCTAGCTAA AGGATTGTAA ATGCACCAAT CAATGCTCAG TGTCTAGCTA
2751 AAGGATTGTA AATGCACCAA TCAGCACTCT GTAAAATTGA CCAATCAGCG
2801 TTCTGTAAAA TGGACCAATC AGTGGTCTGT AAAATGGACC AGTCAGCAGG
2851 ATGTGGGCGG GGCCAAAAAA GGGAATAAAA GCTGGCCACC GCCAGGCTCC
2901 CCACCAGCCT GCAGCGAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 172 bp to 1047 bp; peptide length: 292
 Category: similarity to unknown protein
 Prosite motifs: WW_DOMAIN_1 (19-24)

```

1 MYQGEFGLNM KLGYGKFSWP TGESYHQFY RDHCHGLGTY MWPDGSSFTG
51 TFYLSHREGY GTMYMKTRLF QTHCHNDIVN LLLDCGADV N KCSDEGLTAL
101 SMCFLHLYPA QSFKNVAER TIPEPQEPK FPVVPILSSS FMDTNLESly
151 YEYVNVPSQGS YELRPPPPAPL LLPRVSGSHE GGHFQDTGQC GGSIDHRSSS
201 LKGDSPLVKG SLGHVESGLE DVLGDTDRGS LCSAETKFES NLCVCDFSIE
251 LSQAMLERSA QSHSLKMAS PSPCTSSFDPK GTMRRMALSM IE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_23n16, frame 1

TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds., N = 2, Score = 138, P = 1.1e-06

TREMBL:AF019380_1 product: "putative phosphatidylinositol-4-phosphate
 5-kinase"; Arabidopsis thaliana putative
 phosphatidylinositol-4-phosphate 5-kinase mRNA, complete cds., N = 2,
 Score = 138, P = 1.4e-06

PIR:T02098 probable phosphatidylinositol-4-phosphate 5-kinase -
 Arabidopsis thaliana, N = 2, Score = 135, P = 6.7e-06

>TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds.
 Length = 683

HSPs:

Score = 138 (20.7 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
 Identities = 23/61 (37%), Positives = 35/61 (57%)

```

Query:      1 MYQGEFGLNMKLGYGKFSWPTGESYHQFYRDHCHGLGTYMWPDGSSFTGTGYLSHREGY 60
             MY+G++      G GKFSWP+G +Y G+F      G GT+      DG ++ GT+      + G+
Sbjct:     34 MYEGDWKRGKASGKGFSPSGATYEGEFKSGRMEGFGFTGADGDTYRGTWVADRKHG 93

Query:      61 G 61
             G
Sbjct:     94 G 94

```


Score = 112 (16.8 bits), Expect = 9.7e-04, Sum P(2) = 9.7e-04
Identities = 19/51 (37%), Positives = 27/51 (52%)

Query: 12 LGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGYGT 62
+G GK+ W G Y G + R G G + WP G+++ G F EG+GT
Sbjct: 22 IGSGKYLWKDGCMEGDKRGKASGKGFWSGATYEGEFKSGRMEGFGT 72

Score = 97 (14.6 bits), Expect = 4.4e-02, Sum P(2) = 4.3e-02
Identities = 19/60 (31%), Positives = 32/60 (53%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGYG 61
Y+GEF G+G F+ G++Y G + D HG G + +G + GT+ + ++G G
Sbjct: 58 YEGEFKSGRMEGFGTFTGADGDTYRGTWVADRKHGHGQKRYANGDFYEGTWRRNLQDGRG 117

Score = 93 (14.0 bits), Expect = 1.2e-01, Sum P(2) = 1.1e-01
Identities = 18/62 (29%), Positives = 34/62 (54%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGYG 61
Y+G + + K G+G+ + G+ Y G + R+ G G Y+W +G+ +TG + + G G
Sbjct: 81 YRGTWVADRKHGHGQKRYANGDFYEGTWRRNLQDGRGRYVWRNGNQYTGWRIGVISGK 140

Query: 62 TM 63
+
Sbjct: 141 LL 142

Score = 91 (13.7 bits), Expect = 2.0e-01, Sum P(2) = 1.8e-01
Identities = 18/51 (35%), Positives = 24/51 (47%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGYGT 52
Y GE+ + + G G WP G Y G + G G + W DGSS G +
Sbjct: 127 YTGWRIGVISGKGLLVWPNGNRYEGLWENGIPKGNVFTWSGSSCVGAW 177

Score = 90 (13.5 bits), Expect = 2.6e-01, Sum P(2) = 2.3e-01
Identities = 17/60 (28%), Positives = 31/60 (51%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGYG 61
Y+G + N++ G G++ W G Y G++ G G +WP+G+ + G + +G G
Sbjct: 104 YEGTWRRNLQDGRGRYVWRNGNQYTGWRIGVISGKGLLVWPNGNRYEGLWENGIPKGN 163

Score = 45 (6.8 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
Identities = 14/62 (22%), Positives = 26/62 (41%)

Query: 215 VESGLEDVLGDTDRGSLCSAETKFESNLCVCFD--SIELSQAMLESAQSHSLKMASPS 272
V+SG + G+ +C E+ E+ CD ++E S +R + + +
Sbjct: 205 VDSGAGSLGGEKVFPRIWESDGEAGDITCDIIDNVEASMIYRDRISVDRDGRQFKKN 264

Query: 273 PC 274
PC
Sbjct: 265 PC 266

Pedant information for DKFZphfbr2_23n16, frame 1

Report for DKFZphfbr2_23n16.1

[LENGTH] 292
[MW] 32214.44
[PI] 5.51
[HOMOL] TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for AtPIP5K1,
complete cds. 7e-08
[BLOCKS] BL01137A Hypothetical YBL055c/yjjv family proteins
[PROSITE] WW_DOMAIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] PKC_PHOSPHO_SITE 5
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 4.11 %

SEQ MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLHSHREGY
SEG
PRD cc

SEQ GTMYMKTRLFQTHCHNDIVNLLDCGADVKNKCSDEGLTALSMCFLHYPASFKPNVAER
SEG
PRD cccchhhhhheeeccccchhhhhccccccccccccccccchhhhhhhhhccccccccccccc

SEQ TIPEQEPKFPVVPILSSSFMDTNLESLYEVNVPSSQGSYELRPPAPLPLPRVSGSHE


```

SEG .....XXXXXXXXXXXX.....
PRD eccccccccceeeeeccccccccceeeeeccccccccccccccccccccccccc

SEQ GGHFQDTGQC GGSIDHRSSSLKGD SPLVKGSLGHVESGLE DVLGDTDRGSLCSAETRFES
SEG .....
PRD cccccccccccccccccccccceeeccccccccccccccccccccceeeeecccccc

SEQ NLCVCDFSIELSQAM LERSAQSHSLLKMASPSPCTSSFDKGTMRMRMALS MIE
SEG .....
PRD cccccchhhhhhhhhhhhhhhhhccccccccccccccccchhhhhhhccc

```

Prosites for DKFZphfbr2_23n16.1

PS00005	55->58	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	121->125	CK2_PHOSPHO_SITE	PDOC00006
PS00006	140->144	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00008	45->51	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	177->183	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS01159	19->44	WW_DOMAIN_1	PDOC50020

(No Pfam data available for DKFZphfbr2_23n16.1)

DKFZphfbr2_23o24

group: brain derived

DKFZphfbr2_23o24 encodes a novel 139 amino acid protein with similarity to CAAX-box proteins.

The CAAX box is a prenyl group binding site found in a number of eukaryotic proteins, such as which is found in Ras- and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap, as well as in nuclear lamins A and B, some G protein alpha and gamma subunits and some dnaJ-like proteins. These proteins are posttranslationally modified at this site by the attachment of either a farnesyl or a geranyl-geranyl group to a cysteine residue.

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to lectins

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3564 bp

Poly A stretch at pos. 3541, no polyadenylation signal found

```

1 GAATGGCTCC GCAGATGGCC GGCCTGAGA GCCAGCAAGA AGCGGAGGAG
51 ATGGGCTTC AGCAGGGGGT TCGGGGGGGA GCTTTAACT GAGCCCTGTA
101 AACATGGCAG AACTGCTCAG TGGGAGACTC TCAGCACAGA CGGTTCATGGG
151 GAAGTGAGTG CAGTTCATTT GTAATCTTGT TGTCAGTTC TGGGTTTTTT
201 TTGTTTTGTTT CGTAACTTTA AAGGTATGCA CTTTATATAG ATTTATTTAT
251 TTGCTGGGAC CGTTACTCAG AGTTCCTAGA AATGTACACA GCTTTTTTAC
301 CAGGGTTACT CCTCAGAATC ACTTGTCAC TCTTTAAATG AATGAATGAA
351 TGTGCCAGGC CCTATGCCTG GAGGTGGGA GCTTCATCTA CATCACATTC
401 TAACAGGTGA CCACTGGGGT AAGCACTGTG TGACTGCAAA GCCAGGGTGT
451 GTTCCCATCA ACACCCAGAT GACCGTGCCT ATGTGCCCTT GTTGCTCTCC
501 CTCCAGGACT GCCTCCTCAC CCCACCCCTT TCTGCAGCTC CTCATCTAAA
551 CATCTCGCCT GGTGAGGTCA CGGCTTAGCC TGTTGGCCAG TGGCCCCACC
601 ACCATCCTTC CCCCTGTGCA GATTGGAGGA GGCCAGGTCT CTCCTCTTAG
651 CTCCTATGTC CCCTTCACCC CCCATGGCAC AGATGAGACA TTCACAGAGT
701 TTGAGATGA TGGAAAGAGAA GACTCCAGGT TGCCAGGTGT GTCCACTCTC
751 AGGAACCCCC AGCCCAAGCC TCACTGCTCG TGTTCCACGC CAACCCACGC
801 ACGGGGGATA CGCCGGTGCT GTTTCCTGTC TCAGATACAA CCAGTTACCA
851 GAAACGACCT CACCCCTCCA ACCACTTTCC AAGGTGCCAG GACAGAGAAG
901 CCCTTCACGT GCCCACCCAG GGCAGTTGAC AGAGGGATGC CCTCCTTGGA
951 GGGGAGCCTC ACCTCTACCC ACAGGGCCGC GGCCTTGCTC TGGATTCTCA
1001 CCGGGGAGT CACGTACAGG TGGAGAGGTC CCATGTCAGC CAGTTCCTTG
1051 GTGGGGGTCA TGTAATCTGA AATGACCTGC CGATGGTCCA GGCTGAGCCA
1101 GGGGAGGTGA GCCTGGGTGC CTTTTGTGTG CCTACTCTGA CTTGAGTTGG
1151 ATTCATGCCA CAGACCCACC TTCCTGAGCA ACAACACATA TAGCCACCAA
1201 CACAAGAGCC AGGCACACAC TGAGCAGAGA AAGTCCCTGT CGCCTCACCA
1251 CCCAAAAACT CCAGCTTTGC AGAGACCAAG GTTCTTCTCT ACCTTTGCAG
1301 AAGCCTCTGT GACCAAAACC GGAGCTTGCC CTTCTGAGGC CTCAGCATT
1351 TCTCCAGGTG TTTTTCAGAG GACTTGGTTT AAATTGTGTC ACCCCAAATG
1401 TGGTCTTTCC CGGATCATGA AAGGATCTGC CGCAAAGGTG AATCTGAGTC
1451 TCCTCAGAGT CATATGAGAC TGAAACTGCT TATAACATT CCGTGACCTA
1501 ATAAGTCTTC CAAAAATGTA GGGTATTAAG AGTTTAGTGA CATTAAAAAG
1551 TTTAGTCGAA AATATCGTGA TTCAGGTATA TTTAGACATT TGATTCTATG
1601 CAAATTGCCA CTGTTAACAG AAAACACACC CCAAGCACAT TAATGCCTAG
1651 ATATTTCAAA CCCTTTTCTG CCCACACATT CTTAAAAATA ATATACTGAG
1701 AAATCTATAT ACAGGTTTTT TTTAATTAG CTTGGAAAAG AGCAGTTGTA
1751 TTCTGTTTGA ACAGCTGCTA ATGTCAATTC CTGTGGGAAG AAAGACCAAA
1801 GAACATGGAG TTACACCAAG AATTTTAAAA CAAAGACGCT GTCCCTTTCC
1851 TGAGCACCGT GCAGCCAAGA CTGAGAGATC AGTCTGAGAC CTGTGATTAA
1901 GGAGTGTTTT CTACATAGCG TATAATTATG GAGCCACACA AGTGGGCCAT
1951 TACTCTGTGT AGTGCTTCAT GTTTGAGGTA TTTTCGTGTT CCAACTTACA
2001 TTAAGTGTGT TATAAAACAG GAAAAATCCA CGAGCAGGTA TTGACACTAT
2051 CCATATTAGA TCATCACAAA ATTATATATA TAGCAGAGTC ATAAACAATG
2101 AGAAACGGTC TTCCACACT TGCTTTAAAT GGCCATGACC TAGTGTTTAG
2151 GGAAAGCAGT AAAATCAGCG AGGAGCTCGT GGGAAAAATG AGACGGGCCC
2201 TGAGGGGGTG ACTCATGGGC CAAGCAGGGC CACACAGGTA CCAGGCGGCC
2251 ACGTCTCTC CTGCTCTCA CTCTCTGGAG ACTGGACTTC CTTTACTGCC
2301 TCCTTTCTGA CATTTCCTAG ACATCAGACT TTGCTACTTA GTACACAAAC
2351 GGGTTCCTCT TTTAAATTG TCACTCTAG TTAGCATTGT CAGAAGCTGT
2401 GAAAAATTAC AGAGAGATGA TGTGTTGGGT AAGAGATGGT TTTAAAGTCC

```



```

2451 AGCTTGCTGT TTTTCATTAA GTGCTCTGAA AATGAGTAAG TGGCGTTCCT
2501 GGAGGGGAAC AATCATATAA TTCCGCAGGG TGGGTCTAAA CTTGTTTTCT
2551 GATAGTGTTT AGCAGCTCAT GGCTCTGAGG GCACCTGATA ACACAGCAGC
2601 CAGGCGCTGA TGAGAAGTGT GTGCCAGACA GACCCGAGTG TGGCTTGGCT
2651 CTTGCCTTAT GTTCCTTTCT CTGTTACAGAG AAGCGTGAGA TGAGATTTTG
2701 TGATTATATT GCACTCCTTG GGCTGACTTT CCCATGCACA GAATGTTTTA
2751 CACATCCTGA TAGCTGAGCT GAAAATGCAA AGAGAAGGGA AAATGCCTTA
2801 AATTGTTCTG GCTAATTAG AAGCAGCAGG CCTTGAAGT CTTTGTCTTG
2851 TGTCCCTGAA CAAATCTTAT GGGAGCTCTG GTACCTATGC CAGAAAATGC
2901 ACATAGGCAC AACACTTTTA CATACACGTT CACACACCCC ACCCTTATGG
2951 AGAAGCTTTT TCTAAATAAG AGAAAGAAAA ATTTAAGAC TTACAAGTTA
3001 TGTTTAGGTA TTTTACATGG TTCAGAAAAA AAGACATGAA GCGGTATAAA
3051 CTGAGAAGTC TTGTTCCAC AACCCACCGT GCCAGGTACA CATAACCATT
3101 TTTATTACAC TCTAGCTTGT GCTTCCAATG TTTGTTAGGC ATATGTAAT
3151 AAGTGAATAG ATAAGCATT CTCCCTCCTT TTGCTGACAT GAGTGGTGGC
3201 ATGTTTTGCC CCTGGCTTTT ATCCCTTGAC CCCATTCCAG TACCTAGAGA
3251 CCTGCTTCAT TTTTTTAGAT GTGTAATACT TCATGTGTGC GTGTGCCTTA
3301 GTGATTAAC CGTGCACTGT GCAGGGACAT CGGGCTGGGA TCAGTTTGTT
3351 CACTGATATA TACAGCGCTG CGGGAGATAC CCTCACATGT GTATCATTTG
3401 GTCCATGTGC AGGTGTGTCT GGAAGATAGA ATTCTAGGCG TAGAATTGAT
3451 AGGTTAAATG TATTATAGG GAAAAATCA ATATAAACT TTGCGTGTA
3501 TGATATTTCG GTGCTTTTTT TTTAATTTT TTTACCCAAA TAGTAAAAAA
3551 AAAAAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 656 bp to 1072 bp; peptide length: 139
 Category: similarity to known protein

```

1 MSPSPMAQM RQSLSQMM EKTGQCQVCP LSGTPSPSLT ARVPSQPQHG
51 GYAGAVSLLR YNQLPETTSP LQPLSKVPGQ RSPSLAHPGQ LTEGCPPWRG
101 ASPLPTGFRP CPGFSPGQSR QDGEVPCQPV LWWGSCSLK

```

BLASTP hits

Entry CEEGAP7_1 from database TREMBL:
 gene: "EGAP7.1"; Caenorhabditis elegans cosmid EGAP7.
 Score = 123, P = 2.3e-07, identities = 35/103, positives = 44/103

Entry MMBPC35_1 from database TREMBL:
 Mouse carbohydrate binding protein 35 mRNA, 3' end.
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Entry A28651 from database PIR:
 galactose-specific lectin - mouse >TREMBL:MMAC2A_1 Mouse mRNA for
 Mac-2 antigen
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Alert BLASTP hits for DKFZphfbr2_23o24, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23o24, frame 2

Report for DKFZphfbr2_23o24.2

```

[LENGTH]      139
[MW]           14748.91
[pI]           8.90
[PROSITE]      PRENYLATION 1

```



```

SEQ      QDGEVPCQPVLWWGSCSLK
PRD      CCCCCCCCCCCCCCCCCCCC

```

PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00006	119->123	CK2_PHOSPHO_SITE	PDOC00006
PS00008	50->56	MYRISTYL	PDOC00008
PS00013	126->137	PROKAR_LIPOPROTEIN	PDOC00013
PS00294	136->140	PRENYLATION	PDOC00266

191

DKFZphfbr2_23o5

group: brain derived

DKFZphfbr2_23o5 encodes a novel 360 amino acid protein with no known similarity

No informative BLAST results: no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

potential start at Bp 24 matches Kozak consensus ANNatG

Sequenced by AGOWA

Locus: /map="7q21-q22"

Insert length: 1736 bp

Poly A stretch at pos. 1714, polyadenylation signal at pos. 1680

```

1  GGGGGAGGAT CAAAGTAGGC AAGATGGCGT CGAGCGGCGG GGAGCCAGGG
51 AGTTTATTTG ATCACCACGT CCAGAGGGCG GTATGCGACA CACGGGCCAA
101 ATATCGAGAG GGACGACGGC CTCGTGCTGT GAAGGTATAT ACAATCAATT
151 TGGAACTCA GTACTTATTA ATACAAGGAG TTCCTGCTGT GGGAGTCATG
201 AAGGAATTAG TTGAGCGATT CGCTTTATAT GGTGCAATTG AACAGTACAA
251 TGCTCTACAT GAATACCCAG CAGAAGACTT TACTGAAGTT TATCTTATTA
301 AATTATGAA CTTACAAAGT GCAAGGACAG CCAAGAGAAA AATGGATGAA
351 CAGAGTTTCT TCGGTGGATT GCTTCATGTG TGCTATGCTC CAGAATTGTA
401 AACAGTTGAA GAAACTAGAA AAAAACAACA AATGCGGAAG GCATATGTAG
451 TAAAACTAC TGAAAAATAA GACCATTACG TGACAAAGAA GAAATTGTT
501 ACAGAGCATA AAGACACAGA GGATTTTAGA CAAGACTTCC ACTCAGAGAT
551 GTCTGGATT TGTAAGCTG CTTTGAACAC TTCTGCAGGG AACTCAAATC
601 CTTATCTTCC GTATTCTCTG GAATTGCCTT TATGTTATTT CTCCTCAAAA
651 TGTATGTGTT CATCCGGGGG ACCTGTAGAC AGAGCACCAG ACTCCTCTAA
701 GGATGGTAGA AACCATCATA AAACAATGGG GCATTATAAC CACAATGACT
751 CTTTGCGGAA AACACAGATA AACTCTTTGA AAAACTCAGT GGCCTGCCCT
801 GGTGCACAAA AGGCTATTAC GTCTTCAGAG GCAGTTGACA GATTTATGCC
851 TAGGACAACA CAATGCGAGG AGCGCAAAAG AAGAAGAGAA GATGATCGTA
901 AACTTGGAA TTTTCTTCAA ACAAACCCAA CTGGTAATGA GATTATGATT
951 GGACCTCTGT TACCAGACAT CTCTAAAGTG GATATGCACG ATGACTCATT
1001 GAATACAACG GCGAATTTAA TTCGGCATAA ACTTAAAGAG GTATTTCTATC
1051 TGTGCCAAAG CCTCCAGAGG ACAAGCCAGA AGATGTACAT ACAAGTCATC
1101 CATTAAACA AAGAAGAAGA ATATAGAGTG CCAGCAGCAA CTTAGTATTT
1151 TCTAAAAAGA ACATTTATTA TTTATTTTGA GCCTGTCATT TTAATTCTTC
1201 AAGAGATTTT ACTGCTGGTA TTTTGTGATG CACTCCTCTT TGTAAATTCA
1251 TTCAAGCCAT TTGTCTAAAG TCATTTCTTT GTTTTTTGGG AGATGGAGTC
1301 TTGCTCTGTT GCCCAGGCTG GAATGCAGTG GCGTGATCTC GGCTCACTGC
1351 AACCTCCACC TCCCGGGTTC AAGCGATTCT CCTGCCTCAG CCTCCTGAGT
1401 ATCTGGGATT ACAGGCGTGC ACCACCATGC CTGGCTAAGT TTTGTGTTTT
1451 TTTTAGTAGA GATGGGTTT CACCATATTG GTCAGGCTGG TCTCGAACTC
1501 CTGACCTTGT GATACACCTG CCTCAGCCTC CCAAGGGAT GAGCCACCGC
1551 GCCTGGCCCA TTTCTTCTT TTTTGACCCA TACTTAATGT TGCAGAAATC
1601 ATTCTTGTC TAACATTATC TCTCATGTAC AGTAATTATA TGTAAATTAA
1651 TTGAAGCAA TATGGAACT TTACAATAGA AATAAAGATA GGCAGCCAGC
1701 GTCTGTTTCC AATTATAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry AC005156 from database EMBL:
Homo sapiens PAC clone DJ1099C19 from 7q21-q22, complete sequence.
Score = 2897, P = 2.4e-154, identities = 583/586
2 exons covering Bp 465-1723

Medline entries

No Medline entry

Peptide information for frame 3

 ORF from 24 bp to 1103 bp; peptide length: 360
 Category: similarity to unknown protein

```

1 MASSGGEPGS LFDHHVQRAV CDTRAKYREG RRPRAVKVYT INLESQYLLI
51 QGVPAVGVMK ELVERFALYG AIEQYNALDE YPAEDFTEVY LIKFMNLQSA
101 RTAKRKMDEQ SFFGGLLHVC YAPFETVEE TRKKLQMRKA YVVKTTENKD
151 HYVTKKKLVT EHKDTEDFRQ DFHSEMSGFC KAALNTSAGN SNPYLPYSCE
201 LPLCYFSSKC MCSSGGPVDR APDSSKGRN HHKTMGHYNH NDSLRRKTQIN
251 SLKNSVACPG AQKAITSSEA VDRFMPRTTQ LQERKRRRED DRKLGTFLOT
301 NPTGNEIMIG PLLPDISKVD MHDDSLNTTA NLIRHKLKEV FHLCQSLQRT
351 SQKMYIQVIH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23o5, frame 3

TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence., N = 2, Score = 114, P = 3.6e-11

>TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence.
 Length = 227

HSPs:

Score = 114 (17.1 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 21/41 (51%), Positives = 29/41 (70%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPFETVEETRKKLQMRKAYVV 143
 AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+
 Sbjct: 51 AKRKLDESSFLGNRLQISYAPYENVNDTKDKLESRRKEVL 91

Score = 107 (16.1 bits), Expect = 2.6e-10, Sum P(2) = 2.6e-10
 Identities = 50/191 (26%), Positives = 83/191 (43%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPFETVEETRKKLQMRKAYVVKTTENKDHVTKKKLVTEH 162
 AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+ + T + VT+
 Sbjct: 51 AKRKLDESSFLGNRLQISYAPYENVNDTKDKLESRRKEVLARLNPQKEKSTSQ--VTKL 108

Query: 163 KDTEDFRQDFHSEMSGFC KAALNTSAGNSNPYLPYSCEPLCYFSSKCMCSSGGPVDRAP 222
 + D S + + GN+ P S + YF+S M + V
 Sbjct: 109 AGPALQTQDNVSSQRREMEYQFHR--GNA-PVTRVSSDQE--YFASSSMNQTVKTV---- 159

Query: 223 OSSKGRNHHKTMGHYNHNSLRKTQINSLKNSVACPGAQKAITSSEAVDRFMPRTTQLQ 282
 K + + + +H + ++ N + P +Q S R P ++Q+Q
 Sbjct: 160 -REKLNKTRREENISLSHCKQIEESG-NQKRLQ--PSSQTQPEESGNQKRLQP-SSSIQ 213

Query: 283 -ERKRRREDDRK 293
 + KR R D+R+
 Sbjct: 214 PDLKRTRVDNRR 225

Score = 102 (15.3 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 22/55 (40%), Positives = 38/55 (69%)

Query: 26 KYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMKELVERFALYG AIEQY--NALDE 80
 +Y++ P AV+VYT+ ES+Y++++ VPA+G +L+ F YG +E++ LDE
 Sbjct: 3 RYKD-ETP-AVRVYTVCDSESRMIVRNVPALGCGDDLMRLFMTYGEVEEFAKRLDE 57

Pedant information for DKFZphfbr2_23o5, frame 3

Report for DKFZphfbr2_23o5.3

```

[LENGTH]      360
[MW]           41105.85
[pI]           8.89
[HOMOL]        TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC
F15K20 genomic sequence, complete sequence. 5e-12
[PROSITE]      1
[PROSITE]      MYRISTYL 2
[PROSITE]      CK2_PHOSPHO_SITE 7

```



```

SEQ      MASSGGEPSLFDHHVQRAVCDTRAKYREGRRPRAVKVYTINLESQYLLIQGVPVAVGMVK
SEG      .....
PRD      cccccccccceeeceeeehhhhhhhhhccccceeeeeeccccceeeeeeccccchhh

SEQ      ELVERFALYGAIEQYNALDEYPAEDFTEVYLIKFMNLQSARTAKRMDEQSFSGGLLHVC
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhccccceeeeeeehhhhhhhhhhhhhhhhhccccceeee

SEQ      YAFEFETVEETRKKLQMRKAYVVKTTENKDHVYTKKKLVTEHKDTEDFRQDFHSEMSGFC
SEG      .....
PRD      eccccchhhhhhhhhhhhhhhheeeccccceeeeeeeeeeeccccchhhhhhhhhcccce

SEQ      KAALNTSAGNSNPYLPYSCELPCLFYSSKCMCSSGGPVDRAPOSSKDGRNHHKTMGHYNH
SEG      .....
PRD      eeeccccccccccccccccccccceccccccccccccccccccccccccccccccccccccc

SEQ      NDSLRLKTQINSLKNSVACPGAQKAITSSAEDRFMPRTTQLQERKRRREDDRKLGTFLQT
SEG      .....
PRD      cccceeeccccccccccccccccceeeccccceeeccccchhhhhhhhhhhhhccccceeeeee

SEQ      NPTGNEIMIGPLLPDISKVDMDHDSLNTTANLIRHKLKEVFLCQSLQRTSQKMYIQVIH
SEG      .....
PRD      cccccceeeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhccccchhhhhccc

```

PS000001	185->189	ASN_GLYCOSYLATION	PDOC000001
PS000001	241->245	ASN_GLYCOSYLATION	PDOC000001
PS000001	327->331	ASN_GLYCOSYLATION	PDOC000001
PS000005	99->102	PKC_PHOSPHO_SITE	PDOC000005
PS000005	102->105	PKC_PHOSPHO_SITE	PDOC000005
PS000005	131->134	PKC_PHOSPHO_SITE	PDOC000005
PS000005	154->157	PKC_PHOSPHO_SITE	PDOC000005
PS000005	207->210	PKC_PHOSPHO_SITE	PDOC000005
PS000005	224->227	PKC_PHOSPHO_SITE	PDOC000005
PS000005	243->246	PKC_PHOSPHO_SITE	PDOC000005
PS000005	251->254	PKC_PHOSPHO_SITE	PDOC000005
PS000005	351->354	PKC_PHOSPHO_SITE	PDOC000005
PS000006	4->8	CK2_PHOSPHO_SITE	PDOC000006
PS000006	10->14	CK2_PHOSPHO_SITE	PDOC000006
PS000006	127->131	CK2_PHOSPHO_SITE	PDOC000006
PS000006	224->228	CK2_PHOSPHO_SITE	PDOC000006
PS000006	266->270	CK2_PHOSPHO_SITE	PDOC000006
PS000006	303->307	CK2_PHOSPHO_SITE	PDOC000006
PS000006	317->321	CK2_PHOSPHO_SITE	PDOC000006
PS000008	5->11	MYRISTYL	PDOC000008
PS000008	260->266	MYRISTYL	PDOC000008
PS000009	29->33	AMIDATION	PDOC000009

194

DKFZphfbr2_2a2

group: brain derived

DKFZphfbr2_2a2.3 encodes a novel 167 amino acid protein with weak similarity to human 52K autoantigen Ro/SS-A

The novel protein contains a C3HC4 Zinc finger "RING finger" motive. This domain is probably involved in mediating protein-protein interactions. Proteins containing a RING-finger are: mammalian V(D)J recombination activating protein (RAG1), mouse rpt-1, human rfp, human 52 Kd Ro/SS-A protein and others.

No informative BLAST results; no predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to 52K autoantigen Ro/SS-A - human

complete cDNA, complete cds, few EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1376 bp

Poly A stretch at pos. 1355, polyadenylation⁷ signal at pos. 1340

```

1 GGGGACTCCA AATTAGAAAG GGGACGTCTA GTGGGTTGCC CGGGAGGGGT
51 GCGGGGAGCG GTCCTGGAAA TAATCTGTCC TCTGTCGCCG GGAAGTGGCG
101 AGGTAGTTCC TTCGCGGTGG AGAGACCTGG AATGGCCAAA TATCAAGGTG
151 AAGTTCAAAG TTTGAAACTG GATGATGATT CAGTTATAGA AGGAGTAAGC
201 GACCAAGTAC TTGTGGCAGT TGTGGTCAGT TTCGCTTTGA TTGCTACCCCT
251 GGTATATGCA CTTTTCAGAA ATGTACATCA AAACATTAC CCAGAAAAACC
301 AGGAGCTAGT AAGGGTACTT CGAGAACAGC TTCAAACAGA ACAGGATGCA
351 CCTGCTGCCA CTCGACAGCA GTTCTACACT GACATGTACT GTCCCATCTG
401 CCTGCACCAA GCCTCCTTCC CGGTGGAGAC CAACTGTGGA CATCTTTTTT
451 GTGGTGCCTG CATTATTGCT TACTGGCGAT ATGGTTCATG GCTTGGGGCA
501 ATCAGTTGTC CAATCTGTAG ACAAACGGTA ACCTTACTCC TAACAGTATT
551 TGGTGAAGAT GATCAGTCTC AGGATGTTCT GAGATTGCAT CAGGATATTA
601 ATGATTATAA CCGGAGATTTC TCAGGGCAAC CCTGATCTAT TATGGAGAGA
651 ATTATGGATC TACCCACTTT ACTGAGGCAT GCATTGAGG AAATGTTTTT
701 AGTCGGGGGC CTTTCTGGA TGTTCGCAT CAGGATAATA CTTTGTTTAA
751 TGGGAGCTTT TTTCTATCTT ATATCACCTC TAGATTTTGT ACCTGAAGCC
801 TTGTTTGGA TTCTAGGCTT TCTAGATGAT TTCCTTGTCA TCTTTTATT
851 GCTTATCTAC ATCTCTATTA TGTATCGAGA AGTGATAACC CAAAGGCTAA
901 CTAGATGAAA AAGGAAACAA AACTGAGTTT ACTAGGATAT CTGAGCTAAT
951 GTAGAATATC AAACAGAAAG ACCCATGGCA GTATAAGCA ATGAAGCAAT
1001 GGAGTATTAT CTCACAAATA TAAACCACT ATAAGACAAA CATTTGATTA
1051 TCATTTGACA AATACCTAGG TATAACTGGA ATTTTCATGT TTGAAGTTCT
1101 AATATTAGT TTAGAATTAT AATGATCTAC AGTTGTATCT TGATTCTATG
1151 TTGCTCTGGA AAAATATGGA ATTATATAAA AAGGGATGCT TTTATATATT
1201 TTTCTTTTCC CCAGAATTAC TTAGATTAAT TAGATGTATA GTAAATATT
1251 GTTAAATGTC AGTTTATCCA TCTTATCCTT CTCAGCAGGT ACCTATATGA
1301 TAATATATAG CTGTGAAACT CATCTAAATA TTTTGTTC AATAAATAT
1351 TATATACTAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 132 bp to 632 bp; peptide length: 167
 Category: similarity to known protein
 Classification: unset


```

1 MAKYQGEVQS LKLDSDSVIE GVSDQVLVAV VVSFALIATL VYALFRNVHQ
51 NIHPENQELV RVLREQLQTE QDAPATRQQ FYTDMYCPIC LHQASFPVET
101 NCGHLFCGAC IAYWRYGSW LGAISCFRQ QTVTLTLTVE GEDDQSQOVL
151 RLHODINDYN RRFSGOP

```

No BLASTP hits available

TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; *Caenorhabditis elegans* cosmid Y38F1A, N = 1, Score = 194, P = 2e-15

TREMBLNEW:AB025011_1 gene: "TRIF"; product: "Trif-d"; Mus musculus
mRNA for Trif-d, complete cds., N = 1, Score = 108, P = 2.6e-06

```
>TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A
      Length = 283
```

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 52/149 (34%), Positives = 78/149 (52%)

```

Query:      16 DSVIEGVSQDLVAVVVSFALIATLVYALFRNVHQNIHPENQELVRVLRLEQLQTEQDAPA 75
             D +E ++ Q+ +A+ V F ++ + A Q Q E R Q+ T++
Sbjct:      41 DPDVE-LATQITMAIAVIF-IVKAI FDAWQSSRRQRAASRMEDNAE---RNQIITQRRRI 96

Query:      76 ATROQFYTDMYCICLHQASFPVETNCGHLFCGACIIAYWRYGWSLGA-ISCPICRQTVT 134
             A Q Q + CPICL ASFPV T+CGH+FC CII YW+ + C +CR T
Sbjct:      97 ALHQSSHE---CPICLANASFPVLVTDGHI FPCCECI IQYWQQSKAIVTPCDAMCRSTFY 153

Query:      135 LLLTV----FGEDDQSQDVLRLHQ-DINDYNRRFS 164
             +LL V G +++ D ++ + I+DYNRRFS
Sbjct:      154 MLLPVHWPTMGTSSEETDDHIQENNRIIDYNRRFS 188

```

Pedant information for DKFZphfbr2_2a2, frame 3

Report for DKFZphfbr2_2a2.3

```
[LENGTH]      167
[MW]           18941.65
[pI]           4.91
[HOMOL]        TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A 1e-13

[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR323c] 2e-04
[BLOCKS]       BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE]      ZINC_FINGER_C3HC4 1
[PFAM]         Zinc_finger, C3HC4 type (RING finger)
[KW]           Irregular
[KW]           3D
[KW]           LOW COMPLEXITY 6.59 %
```

```
SEQ      MAKYQGEVQSLKLDSDSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNHIPENQELV
SEG      .....XXXXXXXXXXXX.....
1rmd-    .....

SEQ      RVLREQLQTEQDAPAATRQQFYTDMYCPICLHQAFFPVETNCGHLFCGACIIAYWRYGSW
SEG      .....HHHHHHBTTTTEETTTEEETTTTEEEHHHHH--HHHHH
1rmd-    .....

SEQ      LGAISCPICRQTVTLLLTVFGEDDSQDVLRRLHQDINDYNRRFSQP
```


SEG
 1rmd- HCCB-TTTT.....

Prosite for DKF2phfbr2_2a2.3

PS00518 102->112 ZINC_FINGER_C3HC4 PDOC00449

Pfam for DKF2phfbr2_2a2.3

HMM_NAME Zinc finger, C3HC4 type (RING finger)

HMM *CPICFcTFQlDyPWPfdePmMlPCgHsFCypCIrrW.....CP
 CPIC L+ P++++CGH+FC +CI+ + CP

Query 87 CPIC-----LHQ---ASFPVETNCGHLFCGACIIAYWRYGSWLGAISCP 127

HMM mC*
 +C

Query 128 IC 129

DKFZphfbr2_2b17

group: transmembrane protein

DKFZphfbr2_2b17 encodes a novel 285 amino acid protein with similarity to D. melanogaster 30K protein.

The protein contains 3 transmembrane regions.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to Drosophila hypothetical 30K protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 3

Sequenced by Qiagen

Locus: unknown

Insert length: 1426 bp

Poly A stretch at pos. 1345, polyadenylation signal at pos. 1330

```

1 GGGGGTATTT CCAAGGACTC CAAAGCGAGG CCGGGGACTG AAGGTGTGGG
51 TGTCGAGCCC TCTGGCAGAG GGTTAACCTG GGTCAAATGC ACGGATTCTC
101 ACCTCGTACA GTTACGCTCT CCCGCGGCAC GTCCGCGAGG ACTTGAAATC
151 CTGAGCGCTC AAGTTTGTCC GTAGGTCGAG AGAAGGCCAT GGAGGTGCCG
201 CCACCGGCAC CGCGGAGCTT TCTCTGTAGA GCATTGTGCC TATTTCCTCCG
251 AGTCTTTGCT GCCGAAGCTG TGAAGTCCGA TTCGGAAGTC CTTGAGGAGC
301 GTCAGAAGCG GCTTCCCTAC GTCCGAGAGC CCTATTACCC GGAATCTGGA
351 TGGGACCGCC TCCGGGAGCT GTTTGGCAAA GATGAACAGC AGAGAATTC
401 AAAGGACCTT GCTAATATCT GTAAGACGGC GGCTACAGCA GGCATCATTG
451 GCTGGGTGTA TGGGGGAATA CCAGCTTTTA TTCATGCTAA ACAACAATAC
501 ATTGAGCAGA GCCAGGCAGA AATTTATCAT AACCGGTTTG ATGCTGTGCA
551 ATCTGCACAT CGTGCTGCCA CACGAGGCTT CATTCGTTAT GGCTGGCGCT
601 GGGGTTGGAG AACTGCAGTG TTTGTGACTA TATTCAACAC AGTGAACACT
651 AGTCTGAATG TATACCGAAA TAAAGATGCC TTAAGCCATT TTGTAATTGC
701 AGGAGCTGTC ACGGGAAGTC TTTTATAGAT AAACGTAGGC CTGCGTGGCC
751 TGGTGGCTGG TGGCATAATT GGAGCCTTGC TGGGCACTCC TGTAGGAGGC
801 CTGCTGATGG CATTTAGAAA GTACTCTGGT GAGACTGTTC AGGAAAGAAA
851 ACAGAAGGAT CGAAAGGCAC TCCATGAGCT AAAACTGGAA GAGTGGAAAG
901 GCAGACTACA AGTTACTGAG CACCTCCCTG AGAAAATTGA AAGTAGTTTA
951 CAGGAAGATG AACCTGAGAA TGATGCTAAG AAAATTGAAG CACTGCTAAA
1001 CCTTCCTAGA AACCTTCAG TAATAGATAA ACAAGACAAG GACTGAAAGT
1051 GCTCTGAAC TGAAACTCAC TGGAGAGCTG AAGGGAGCTG CCATGTCCGA
1101 TGAATGCCAA CAGACAGGCC ACTCTTTGGT CAGCCTGCTG ACAAAATTTAA
1151 GTGCTGGTAC CTGTGGTGGC AGTGGCTTGC TCTTGTCTTT TTCTTTTCTT
1201 TTTAACTAAG AATGGGGCTG TTGTACTCTC ACTTTACTTA TCCTTAAATT
1251 TAAATACATA CTTATGTTTG TATTAATCTA TCAATATATG CATACATGAA
1301 TATATCCACC CACCTAGATT TTAAGCAGTA AATAAACATC TTCGCAAAAG
1351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HSG19630 from database EMBL:

human STS A001T27.

Score = 961, P = 1.2e-36, identities = 193/194

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 189 bp to 1043 bp; peptide length: 285
Category: similarity to unknown protein

BLASTP hits

Alert BLASTP hits for DKFZphfbr2 2b17, frame 3

HSPs:

```

Query:      30 ADSEVLEERQKRLPYVPEPPYPYPSGWDRLRELFGKDEQQRISKDLANICKTAAATAGIIGW 89
            AD V +E + ++ E+G +RL+++F DE I +L ++ + +IG
Sbjct:      23 ADEIVDKENKTYKAFLASKPPEETGLERLQKMFITIDFGSIFSELNSVYQAGFLGLFLIGA 82

Query:      90 VYGGIPAFIHAQQYIEQSAEIIYHNRFDAVQSAHRAATRGFIYGRWRGWRTAVFVTIF 149
            +YGG+ A ++E +QA ++ + FDA + T F +G+WGWR +F T +
Sbjct:      83 IYGGVTVQSRVAYMFMENNQATAFKSHFDAKKKLQDQFTVNFAGKGFKGWVRVGLFTTSY 142

Query:      150 NTVNTSLNVYRNKDALSHFVIAGAVTGSFLFRINVGLRGLVAGGIIGALLGTPVGGLLMAF 209
            + T ++VYR K ++ ++ AG++TGS++++GLRG+ AGGIIG LG G +
Sbjct:      143 FGIITCMSVYRGKSSIYEYLAAGSITGSLYKVSLGLRMAAGGIIGGFLGGVAGVTSLLL 202

Query:      210 QKYSGETVQERKQKDRKALHELKLEEWKGRQLVTEHLPEKIESSLQEDEPE 260
            K SG +++E ++ ++K RL E++ + + ++ PE
Sbjct:      203 MKASGTSME-----VRYWYQWRLLDRDENIQAFKKLTEDENPE 242

```

Pedant information for DKF2phfbr2 2b17, frame 3

Report for DKFZphfbr2 2b17.3

```
[LENGTH]          285
[MW]               32177.88
[pI]               8.65
[HOMOL]            PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila
melanogaster) 7e-20
[PROSITE]          MYRISTYL          7
[PROSITE]          CK2_PHOSPHO_SITE      5
[PROSITE]          ASN_GLYCOSYLATION     1
[KW]               SIGNAL_PEPTIDE 25
[KW]               TRANSMEMBRANE 3
[KW]               LOW_COMPLEXITY        5.96 %
```

```

SEQ      MEVPPPPAPRSFLCRALCLFPRVFAAEAVTADSEVLERQKRLPYVPEPPYPESGWRDLRE
SEG
PRD      CCCCCCCCCcccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      LFGKDEQORISKDLANICKTAATAGIIGWVYGGIPAFIHAQQQYIEQSQAIEYHNRFDAV
SEG
PRD      hhccccchhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      QSAHRAATRGRFIRYGRWRGWRTAVFTTIFNTVNTSLNVYRNKDALSHEFVIAGAVTGSLFR
SEG
PRD      hhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....M

SEQ      INVGLRGLVAGGIIGALLGTPVGGLLMAFQKYSGETVOERKQKDRKALHELKLEEWKGRLL

```



```

SEG      ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

```

SEQ      QVTEHLPEKIESSLQEDPENDAKKIEALLNLPRNPVIDKQDKD
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphfbr2_2b17.3

PS00001	153->157	ASN_GLYCOSYLATION	PDOC00001
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2b17.3)

DKFZphfbr2_2b5

group: cell structure and motility

DKFZphfbr2_2b5 encodes a novel 957 amino acid protein with strong similarity to collagens.

The novel protein contains the typical (xxG)n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain.

The new protein can find application in modulation of connective tissue, bone and cartilage development and maintenance.

similarity to collagen proteins

shows typical (xxG)n repeat of collagen proteins
[PFAM] von Willebrand factor type A domain

Sequenced by Qiagen

Locus: /map="6"

Insert length: 4160 bp

Poly A stretch at pos. 4141, polyadenylation signal at pos. 4119

```

1  GGGGGCCCCG  TGCAGGGAGA  ACGGACTCCG  GCGGGAGGGC  AGCCAATCCG
51  TTTAGCGCGA  GGTCTTGCTC  GGGTTGGGCT  TGCCACTGCC  TGGAACATAC
101 CTGTCCCCCT  GCGCAACAC  TCAGCTGGCT  GCGACCGCAA  CCCCAGCCT
151 GGACACTGCG  CCAGGAATCC  TAAACCAAAA  ATATTAGAAC  GAAACAGAA
201 ACATGGCTCA  CTATATTACA  TTTCTCTGCA  TGGTTTGGT  GCTGCTTCTT
251 CAGAATTCGT  TGTTAGCTGA  AGATGGGGAA  GTAAGATCAA  GTTGCTGTAC
301 TGCTCCGACA  GATTTAGTTT  TCATCTTAGA  TGGCTCTTAT  AGTGTGGCC
351 CAGAAAACCT  TGAATAGTG  AAAAAAGTGG  TTGTCAATAT  CACAAAAAAC
401 TTTGACATAG  GGCCGAAGTT  TATTCAGTT  GGAGTGGTTC  AATATAGTGA
451 CTACCTGTG  CTGGAGATTC  CTCTCGGAAG  CTATGATTCA  GGAGAACATT
501 TGACGCGCAG  AGTGGAAATC  ATACTCTACT  TAGGAGGAAA  CACAAAGACA
551 GGAAGGCCA  TCCAGTTTGC  GCTCGATTAC  CTTTTTGACA  AGTCTCAGC
601 ATTTCTGACT  AAGATAGCAG  TGGTACTTAC  GGATGGCAAG  TCCCAAGATG
651 ACGTCAAGGA  TGCAGCTCAA  GCAGCAAGAG  ATAGTAAGAT  AACATTATTT
701 GCTATTGGTG  TTGGTTTACA  AACAGAAGAT  GCCGAACTTA  GAGCTATTGC
751 CAACAAGCCT  TCGTCTACTT  ATGTGTTTAA  TGTGGAAGAC  TATATTGCAA
801 TATCCAAAT  AAGGGAAGTG  ATGAAGCAGA  AACTTTGTGA  AGAATCTGTC
851 TGTCCAAAC  GAATTCAGT  GGCAGCTCGT  GATGAAAGGG  GATTGATAT
901 TCTTTTGGGT  TTAGATGTAA  ATAAAAAGGT  TAAGAAAAGA  ATACAGCTTT
951 CACCAAAAA  GATAAAGGA  TATGAAGTAA  CATCAAAGT  TGATTATCA
1001 GAACACAAA  GCAATGTTT  CCCAGAAGGT  CTCTCTCCAT  CATATGTATT
1051 TGTGTCTACT  CAAAGATTTA  AAGTCAAGAA  AATTGGGAT  TTATGGAGAA
1101 TATTAATAT  TGATGGAAG  CCACAAATAG  CAGTTACCTT  AAATGGTGTG
1151 GACAAAATCT  TATTATTAC  AACCAACAGC  GTAATTAATG  GCTCACAAGT
1201 GGTTACCTTT  GCTAACCTC  AAGTTAAGAC  GTTGTGATG  GAAGGCTGGC
1251 ACCAAATTCG  TCTCTTAGTA  ACAGAACAA  ATGTGACTTT  GTATATTGAT
1301 GACCAACAAA  TTGAAAACAA  GCCCTTACAT  CCAGTTTAG  GGATCTTGAT
1351 CAATGGGCAA  ACCCAATTG  GAAAATATTC  TGGAAAAGAA  GAACTGTTC
1401 AGTTTGATGT  CCAAAAGTTG  CGAATCTACT  GTGACCCAGA  ACAGAACAC
1451 CGGGAGACAG  CATGTGAGAT  TCCTGGATT  AATGGAGAGT  GCCTTAATGG
1501 TCCCAGTGAT  GTAGGTTCAA  CTCCAGCTCC  CTGTATTGT  CCTCCGGGAA
1551 AACCAGGACT  TCAAGGCCCC  AAAGGTGACC  CTGGACTGCC  TGGGAACCC
1601 GGCTACCCTG  GACAACCTGG  TCAAGATGGT  AAGCCTGGAT  ATCAGGGAAT
1651 TGCAGGGACA  CCAGGTGTTT  CAGGATCTCC  AGGAATACAA  GGAGCTCGAG
1701 GACTACCAGG  TTACAAAGGA  GAACCAAGGC  GAGATGGTGA  CAAGGCTGAT
1751 CGTGGACTTC  CTGGTTTCC  TGGGCTTCAT  GGCATGCCAG  GATCAAAGGG
1801 TGAATGGGT  GCCAAAGGAG  ACAAGGATC  ACCTGGATT  TATGGCAAAA
1851 AGGGTGCAAA  AGGTGAAAAG  GGGAAATGCTG  GCTTCCCTGG  CCTCCCTGGA
1901 CCTGCTGGAG  AACCAGGAAG  ACATGGAAAG  GATGGATTAA  TGGGTAGTCC
1951 CGGTTTCAAG  GGAGAAGCAG  GATCCCTGG  TGCTCCGGGG  CAGGATGGAA
2001 CACGGGGAGA  GCCTGGAATC  CCAGGATTTC  CTGGAAACCG  AGGATTAATG
2051 GGCCAAAAGG  GAGAAATTGG  GCCTCCAGGA  CAGCAAGGAA  AAAAAGGAGC
2101 CCCAGGGATG  CCTGGTTTAA  TGGGAAGCAA  TGGCTCACA  GGCCAGCCTG
2151 GAACACCGG  ATCTAAGGGA  AGCAAAGGTG  AACCTGGAAT  TCAAGGGATG
2201 CCTGGGGCTT  CAGGGCTCAA  GGGAGAACCA  GGAGCAACGG  GTTCCCAGG
2251 AGAACCAGGA  TACATGGGTT  TACCCGGGAT  TCAAGGAAAA  AAGGGGGACA
2301 AAGGAAATCA  AGGTGAAAA  GGTATTACAG  GTCAAAAGGG  AGAAAATGGA
2351 AGACAGGGAA  TTCCAGGGCA  ACAGGGAATT  CAAGGCCATC  ATGGTGCAAA
2401 AGGAGAGAGA  GGTGAAAAGG  GAGAACCTGG  TGTCCGAGGT  GCCATTGGAT
2451 CAAAAGGAGA  ATCTGGGGTG  GATGGCTTGA  TGGGGCCCCG  AGGTCCTAAG
2501 GGGCAACCTG  GGGATCCAGG  TCCTCAGGGA  CCCCAGGTT  TGGATGGGAA
2551 GCCCGGAAGA  GAGTTTTCAG  AACATTTAT  TCGACAAGTT  TGCACAGATG

```



```

2601 TAATAAGAGC CCAGCTACCA GTCTTACTTC AGAGTGGAG AATTAGAAAT
2651 TGTGATCATT GCCTGTCCCA ACATGGCTCC CCGGGTATTC CTGGGCCACC
2701 TGGTCCGATA GGCCAGAGG GTCCAGAGG ATTACCTGGT TTGCCAGGAA
2751 GAGATGGTGT TCCTGGATTA GTGGGTGTCC CTGGACGTCC AGGTGTCAGA
2801 GGATTAAAAA GCCTACCAGG AAGAAATGGG GAAAAAGGGA GCCAAGGGTT
2851 TGGGTATCCT GGAGAACAA GTCTCTCTGG TCCCCAGGT CCAGAGGGCC
2901 CTCCTGGAAT AAGCAAAGAA GGTCTCCAG GAGACCCAGG TCTCCCTGGC
2951 AAAGATGGAG ACCATGGAAA ACCTGGAATC CAAGGGCAAC CAGGCCCCCC
3001 AGGCATCTGC GACCCATCAC TATGTTTATG TGAATTGCC AGAAGAGATC
3051 CGTTCAGAAA AGGACCAAAC TATTAGTGTC TGATGCCTCA TTCAGCAGCC
3101 TAGGCATGGT GCTTTTCTG TGGTCTTTG CATCTCAGGA AGATAACCAA
3151 CAGTATCCCT TGAAGAGAAA CTTAAGTACC TCGGTGTTT TATTTTTTTT
3201 TTCTTATGGA AAAAAATATA AAAGATCACA TATACTGATT TTAAAGGCTC
3251 CTCAGTCATT TGGAGCCCTT GGATTAGCAG CATTAAATTA ATCTCAAGGG
3301 TTTCTTGTA AGTCCATTTA TGTTAATCAA AGTTGAATAT AAAAATCCAC
3351 CATTGCCTGT TAGCCAGTCA GTTTAGTCA CTGTGAAATA TTTCACATTC
3401 AGCCTCCATG CAGTAGAGAT TTGAGTTTAA TTTCATGTCC ATGTGACTTT
3451 CATGTTTCTC ATCTCATAGC TCATGCTACT ACATAAGCCA AAACATGTAT
3501 CTCATCATTG GAAGTAAGAT CAGGGCTGAT ATTCACTGG GATAGACAGT
3551 ATTGGTGAAC TACTCATTTA CTACAGTGTC TCAGCCTTGA TAAAGGGCAG
3601 TGGATTGCCT GTTGTTCGGT GTTGTGAATA GCACCTCTGA ATAAGATTAG
3651 AGTGTTCCTT AATTCATTTT AAACCTCTAA ATTAGATTAA TGGTGGTGCT
3701 AAGAAAGAGT ATTAATTACT TTGGGAATGG TCAAAATTA CATTAAAAAC
3751 ATTTTAGACA AAAAGTTTCA TTGTACATTC AAAGAAAATG TAAGTTTGGA
3801 AGTACTAAAA GACTATTTTA TACTTGTGTA TTAATCGGAA TGTTGTTGT
3851 ATGCCTTCAT TTCCATTTT ACTTATATGT GCATGTCCAT ATATGTTAAT
3901 TTTTATTGTA GCAAGCTAA TGGAAATAAA GCTAATGCTC TAGTTGAAAG
3951 AAAAAGGAAA CTCCTGAAAT CCTAGAATGT CTTGTTATTT TTAGCTGACT
4001 GTAAATATT ATGAACAGTC TTTGTGTATT GTGCTTAATG CTTTGTAAAG
4051 AAACAGAATT TGAATATTT CATCCTTGTC ATGCTCAAAA TTTTGTACA
4101 TGCTTGTTAT TCAGAGTATA ATAAAGTTT GTACAGGCCT GAAAAAATA
4151 AAAAAAATA

```

BLAST Results

Entry HS682J15 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 682J15
 Score = 6240, P = 0.0e+00, identities = 1256/1263
 13 exons matching Bp 2015-4118

Entry HS708F5 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 708F5
 Score = 2775, P = 1.0e-221, identities = 739/912
 10 exons matching Bp 5-1745

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 203 bp to 3073 bp; peptide length: 957
 Category: similarity to known protein

```

1 MAHYITFLCM VLVLLQNSV LAEDGEVRSS CRTAPTDLVF ILDGYSVGP
51 ENFEIVKKWL VNITKNFDIG PKFIQVGVVQ YSDYPVLEIP LGSYDSGEHL
101 TAAVESILYL GGNTKTGKAI QFALDYLFDK SSRFLTIAV VLTGKSQDD
151 VKDAAQAARD SKITLFAIGV GSETEDAELR AIANKPSSTY VFYVEDYIAI
201 SKIREVMKQK LCEESVCPTP IPVAARDERG FDILLGLDVN KVKKKRIQLS
251 PKKIKGYEVT SKVDLSELT NVFPEGLPPS YVFVSTQRFK VKKIWDLWRI
301 LTIDGRPQIA VTLLNGVDKIL LFTTTSVING SQVVTFANPQ VKTLEDEGWH
351 QIRLLVTEQD VTLYIDDOQI ENKPLHPVLG ILINGQTQIG KYSGKEETVQ
401 FDVQKLRIYC DPEQNNRETA CEIPGFNGEC LNGPSDVGST PAPCICPPGK
451 PGLQGPKGDP GLPGNPGYPG QPGQDGKPGY QGIAGTPGVP GSPGIQGARG
501 LPGYKGEFGR DGDGKDRGLP GFPGLHGMFG SKGEMKAGD KGSPPGYGKK
551 GAKGEKGNAG FPGLPGPAGE PGRHGKDGML GSPGFKGEAG SPGAPGQDGT
601 RGEPPGIPGF GNRGLMGQKG EIGPPGQQKG KGAPGMPGLM GSNPSGQPG
651 TPGSKGSKGE PGIQGMFGAS GLKGEFGATG SPGEPMYGL PGIQGGKGDG
701 GNQGEKGIQG QKGENRGQGI PGQGGIQGHH GAKGERGEKG EPGVRGAIGS
751 KGESGVDCGL GPAGPKGQPG DPGPQGPPLG DGKPGREFSE QFIRQVCTDV
801 IRAQLPVLLQ SGRIKNDHC LSQHGSPGIP GPPGPIGPEG PRGLPGLPGR

```


851 DGVPGVLGVGP GRPGVRGLKG LPGRNGEKGS QGFGYPGEQG PPGPPGPEGP
 901 PGISKEGPPG DPGLPGKDGD HGKPGIQGQP GPPGICDPSL CFSVIARRDP
 951 FRKGPNY

BLASTP hits

Entry HSCOL7A1X_1 from database TREMBL:
 gene: "COL7A1"; product: "collagen type VII"; Homo sapiens (clones:
 CW52-2, CW27-6, CW15-2, CW26-5, 11-67) collagen type VII intergenic
 region and (COL7A1) gene, complete cds.
 Score = 949, P = 3.4e-122, identities = 237/553, positives = 281/553

Entry CA17 HUMAN from database SWISSPROT:
 COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC
 COLLAGEN). >TREMBL:HSCOL7A1_1 gene: "COL7A1"; product: "alpha-1 type
 VII collagen"; Human alpha-1 type VII collagen (COL7A1) mRNA, complete
 cds.
 Score = 949, P = 3.6e-122, identities = 237/553, positives = 281/553

Alert BLASTP hits for DKFZphfbr2_2b5, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2b5, frame 2

Report for DKFZphfbr2_2b5.2

[LENGTH] 957
 [MW] 99413.38
 [pI] 8.49
 [HOMOL] PIR:A40020 collagen alpha 1(XII) chain precursor - chicken 9e-90
 [BLOCKS] BL01119B Copper-fist domain proteins
 [BLOCKS] BL00313B
 [BLOCKS] BL01113A C1q domain proteins
 [BLOCKS] BL00420A Speract receptor repeat proteins domain proteins
 [SCOP] dlzoo_ 3.45.1.1.1 Integrin CD11a/CD18 (LFA-1) [Human (Hom 2e-58
 [SCOP] dlido_ 3.45.1.1.2 Integrin CR3 (CD11b/CD18), alpha subunit [Huma 8e-62
 [EC] 3.1.1.7 Acetylcholinesterase 7e-24
 [PIRKW] blocked amino end 1e-43
 [PIRKW] duplication 7e-46
 [PIRKW] cornea 1e-35
 [PIRKW] lung 2e-40
 [PIRKW] leukocyte 1e-42
 [PIRKW] skin 1e-40
 [PIRKW] transmembrane protein 1e-37
 [PIRKW] cartilage 3e-59
 [PIRKW] hydroxylysine 4e-62
 [PIRKW] connective tissue 3e-43
 [PIRKW] triple helix 5e-82
 [PIRKW] homotrimer 2e-37
 [PIRKW] bone 6e-40
 [PIRKW] Alport syndrome 1e-42
 [PIRKW] laminin binding 2e-40
 [PIRKW] liver 2e-40
 [PIRKW] glycoprotein 5e-82
 [PIRKW] carboxylic ester hydrolase 7e-24
 [PIRKW] disulfide bond 7e-46
 [PIRKW] cell binding 7e-46
 [PIRKW] heterotrimer 4e-62
 [PIRKW] calcium binding 8e-28
 [PIRKW] alternative splicing 5e-82
 [PIRKW] coiled coil 5e-82
 [PIRKW] basement membrane 7e-46
 [PIRKW] trimer 5e-82
 [PIRKW] pyroglutamic acid 3e-43
 [PIRKW] hydroxyproline 4e-62
 [PIRKW] extracellular matrix 5e-82
 [PIRKW] chondroitin sulfate proteoglycan 6e-41
 [PIRKW] sulfoprotein 7e-39
 [PIRKW] kidney 1e-42
 [PIRKW] angiogenesis inhibitor 6e-36
 [PIRKW] Ehlers-Danlos syndrome 2e-40
 [SUPFAM] fibronectin type III repeat homology 5e-82
 [SUPFAM] scavenger receptor cysteine-rich domain homology 1e-37
 [SUPFAM] C-type lectin homology 6e-30
 [SUPFAM] collagen alpha 2(I) chain 5e-40
 [SUPFAM] collagen alpha 1(I) chain 6e-44


```

[SUPFAM] fibrillar collagen carboxyl-terminal homology 6e-44
[SUPFAM] animal Kunitz-type proteinase inhibitor homology 2e-38
[SUPFAM] fibronectin type II repeat homology 6e-21
[SUPFAM] complement C1q carboxyl-terminal homology 1e-38
[SUPFAM] collagen alpha 3(VI) chain 2e-31
[SUPFAM] collagen alpha 1(IV) chain 7e-46
[SUPFAM] collagen alpha 1(VI) chain 2e-37
[SUPFAM] von Willebrand factor type C repeat homology 6e-44
[SUPFAM] unassigned collagens 4e-62
[SUPFAM] von Willebrand factor type A repeat homology 5e-82
[SUPFAM] collagen alpha 1(XIV) chain 5e-82
[SUPFAM] pulmonary surfactant protein D 6e-30
[SUPFAM] collagen alpha 1(V) chain 7e-39
[SUPFAM] collagen alpha 1(VIII) chain 1e-38
[SUPFAM] EGF homology 1e-35
[PROSITE] AMIDATION 3
[PROSITE] MYRISTYL 14
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] von Willebrand factor type A domain
[KW] Irregular
[KW] 3D
[KW] SIGNAL_PEPTIDE 23
[KW] LOW_COMPLEXITY 24.24 %

```

```

SEQ MAHYITFLCMVLVLLQNSVLAEDGEVRSSCRTAPTDLVFIIDGSYSVGPENFEIVKKWL
SEG .....
latzB .....CCCEEEEEEECCCCCHHHHHHHHHHH

SEQ VNITKNFDIGPKFIQGVVQYSDYPVLEIPLGSYDSGEHLTAAVESILYLGGNTKTGKAI
SEG .....
latzB HHHHHHCCBTTTTEEEEEEEETTTEEEETTTHHHHHHHHHHCCCCCCCCCHHHHH

SEQ QFALDYLFDKSSRFLTKIAVVLTGKSQDDVKDAAQAARDSKITLFAIGVGSETEDAELR
SEG .....
latzB HHHHHHHHCCTTTTTEEEEEEECCCTTTTHHHHHHHHHHHCEEEEEEECCCCCHHHHH

SEQ AIANKPSSTYVFYVEDYIAISKIREVMKQKLCEESVCPTIPVAARDERGFIDILLGLDVN
SEG .....
latzB HHHGGGGGGGCECCHHHHHHHHHCHHHHHHHH.....

SEQ KKVKKRIQLSPKKIKGYEVTSKVDLSELTSNVFPPEGLPPSYVVFSTQRFVKKIIDLWRI
SEG .....
latzB .....

SEQ LTIDGRPQIAVTLNGVDKILLFTTTSVINGSQVVFANPQVKTLFDEGWHQIRLLVTEQD
SEG .....
latzB .....

SEQ VTLYIDDDQIENKPLHPVLGILINGQTQIGKYSKKEETVQFDVQKLRIYCDPEQNNRETA
SEG .....
latzB .....

SEQ CEIPGFNGECLNGPSDVGSTPAPCICPPGKPGQPGKDPGLPGNPGYPGQPGQDGKPGY
SEG .....
latzB .....

SEQ QGIAGTPGVPGSPGIQGARGLPKYKGEPRDGDGDRGLPGFPGHGMPSKSGEMGAKGD
SEG ..
latzB .....

SEQ KGSPGFYKKGAKGEKGNAGFPGLPGPAGEPRHKGDKLGMSPGFKEAGSPGAPGQDGT
SEG .....
latzB .....

SEQ RGEPIPGFPGNRGLMGQKGEIGPPQGGKKGAPGMPGLMGSNGSPGQPGTPGSKGSKGE
SEG .....
latzB .....

SEQ PGIQGMPSASGLKGEPGATGSPGEPGYMGLPGIQGKKGDKGNQGEKGIQGGKGENGRQGI
SEG .....
latzB .....

SEQ PGQQGIQGHGAKGERGERGEPGVRGAIGSKGESGVDGLMPAGPKGQPGDPGPGPPGL
SEG .....
latzB .....

SEQ DGKPGREFSEQFIRQVCTDVIRAQLPVLLQSGRIRNCDHCLSQHSGSPGIPGPPGPIGPEG
SEG .....

```



```

latzB .....
SEQ   PRGLPGLPGRDGVPLVGVPGRPVGRGLKGLPGRNGEKGSGQFGYPGEQGPFGPPGPEGP
SEG   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
latzB .....

SEQ   PGISKEGPPGDPGLPGKGDGHGKPGIQGQPGPPGICDPSLCFSVIARRDPFRKGPNY
SEG   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
latzB .....

```

Prosites for DKFZphfbr2_2b5.2

PS00001	62->66	ASN_GLYCOSYLATION	PDOC00001
PS00001	329->333	ASN_GLYCOSYLATION	PDOC00001
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	286->289	PKC_PHOSPHO_SITE	PDOC00005
PS00005	393->396	PKC_PHOSPHO_SITE	PDOC00005
PS00005	811->814	PKC_PHOSPHO_SITE	PDOC00005
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	261->265	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	357->361	CK2_PHOSPHO_SITE	PDOC00006
PS00006	393->397	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00006	531->535	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	657->661	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	750->754	CK2_PHOSPHO_SITE	PDOC00006
PS00006	754->758	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	112->118	MYRISTYL	PDOC00008
PS00008	236->242	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	527->533	MYRISTYL	PDOC00008
PS00008	596->602	MYRISTYL	PDOC00008
PS00008	638->644	MYRISTYL	PDOC00008
PS00008	650->656	MYRISTYL	PDOC00008
PS00008	653->659	MYRISTYL	PDOC00008
PS00008	665->671	MYRISTYL	PDOC00008
PS00008	743->749	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00009	547->551	AMIDATION	PDOC00009
PS00009	628->632	AMIDATION	PDOC00009
PS00009	694->698	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_2b5.2

HMM_NAME	von Willebrand factor type A domain		
HMM	*DIVFLIDGSdSIGpNFNRMKDFieRMERMDIGPDwIRVGVVQYSdNP		
Query	37	DLVFILDGYSVGPENFEIVKKWLVNITKNFDIGPKFIQGVVQYSDYP	85
HMM	RqEmrFmFNDYQNKeEILQaIqqMMYwMgggTNTGeAIQYVvrNMfweer		
Query	86	VLE--IPLGSYDSGEHLTAIVESIL-YLGGNTKTGKAIQFALDYLFDKSS	132
HMM	GmRwenvPQVMIIITDGRSQDDIRDPIneMrmaGIqvFaIGIGNhDnNn		
Query	133	RF----LTKIAVVLTDGKSQDDVKDAAQAARD-SKITLFAIGVGSETE--	175
HMM	WeELReIASepdEdHVfYvDdFeeLdnMqeQL*		
Query	176	DAELRAIANKPSSTYVFYVEDYIAISKIREVM	207

DKFZphfbr2_2c1

group: brain derived

DKFZphfbr2_2c1 encodes a novel 697 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 3973 bp

Poly A stretch at pos. 3914, polyadenylation signal at pos. 3900

```
1 GGGGGGATTT CGGCGGCGGA AACATGGCGG TCGCGGCCGG GCCGGTAACG
51 GAGAAAGTTT ACGCCGACAC TGGCCTGTAT TAGCGCGTAT GGCCCTCGGC
101 CCTCGTTCCC CAAGGCGTGC CGCCTCCCTG TTCTCAGTCG CAGGCTGAAG
151 CCTTGTCTGC TCTCCTCCTT TTTGGTTTGG TTTTGGAACT GACTCCGAGG
201 GTTGGGAGAG CGCGTTGGTG GCGACGGCCG AGTCAGATCA CTATAACAA
251 AATTTCACA AGAGAAATG TTGAAATAGG AGTTGCGGAT ACATTGGATA
301 TACTGGATGA AATACAAGCG GTTAATTTT GTAACGTGAG GGAAAGCCC
351 ACATTGCTGG TTACATGTGT AAATCACTGC GTTATTGCTT TAGTCATTGT
401 CTCTATTTAG CAATGACAAG ACTGGAAGAA GTAAATAGAG AAGTGAACAT
451 GCATTCTTCA GTGCGGTATC TTGGCTATTT AGCCAGAATC AATTTATTGG
501 TTGCTATATG CTTAGGTCTA TACGTAAGAT GGGAAAAAAC AGCAAATTCC
551 TTAATTTTGG TAATTTTAT TCTTGGTCTT TTTGTTCTTG GAATCGCCAG
601 CATACTCTAT TACTATTTT CAATGGAAGC AGCAAGTTTA AGTCTCTCCA
651 ATCTTTGGTT TGGATTCTTG CTTGGCCTCC TATGTTTCTT TGATAATTCA
701 TCCTTTAAAA ATGATGTAAA AGAAGAATCA ACCAAATATT TGCTTCTAAC
751 ATCCATAGTG TTAAGGATAT TGTGCTCTCT GGTGGAGAGA ATTTCTGGCT
801 ATGTCCGTCA TCGGCCCACT TTAACAACCA CAGTTGAATT TCTGGAGCTT
851 GTTGGATTGG CCATTGCCAG CACAACATAG TTGGTGGAGA AGTCTCTGAG
901 TGTCAATTTG CTTGTTGTAG CTCTGGCTAT GCTGATTATT GATCTGAGAA
951 TGAATCTTTT CTTAGCTATT CCAAACCTAG TTAATTTTGC AGTTTTGTTA
1001 TTTTTTCTCT CATTGGAAC TCCCAAAAT CCGATTGCTT TTGCGTGT
1051 TTTTATTTGC CTGATAACTG ATCCTTTCTT TGACATTTAT TTTAGTGGAC
1101 TTTTCAGTAA TGAAAGATGG AAACCCCTTT TGTACCGTGG AAGAAATTTG
1151 AGAAGACTTT CAGTCGTTT TGCTGGAATG ATTGAGCTTA CATTTTTAT
1201 TCTTTCGCA TTCAAACCTA GAGACACTCA CCTCTGGTAT TTTGTAATAC
1251 CTGGCTTTTC CATTTTTGA ATTTTCAGGA TGATTTGTCA TATTATTTT
1301 CTTTAACTC TTTGGGATT CCATACCAA TTAATGACT GCCATAAGT
1351 ATATTTTACT CACAGGACAG ATTACAATAG CCTTGATAGA ATCATGGCAT
1401 CCAAGGGATG GCGCCATTTT TGCTTGATTT CAGAGCAGTT GGTGTTCTTT
1451 AGTCTTCTTG CAACAGCGAT TTTGGGAGCA GTTCTCTGGC AGCCAACAA
1501 TGGAATTTTC TTGAGCATGT TCCTAATCGT TTTGCCATTG GAATCCATGG
1551 CTATGGGGCT CTTCCATGAA TTGGGTAAC TTTTAGGAGG AACATCTGTT
1601 GGATATGCTA TTGTGATTCC CACCAACTTC TGCAGTCTCG ATGGTCAGCC
1651 AACACTGCTT CCCCAGAAC ATGTACAGGA GTTAAATTTG AGGTCTACTG
1701 GCATGCTCAA TGCTATCCAA AGATTTTGTG CATATCATAT GATTGAGACC
1751 TATGGATGTG ACTATTCCAC AAGTGGACTG TCATTTGATA CTCTGCATTC
1801 CAAACTAAAA GCTTTCCTCG AACTTCGGAC AGTGGATGGA CCCAGACATG
1851 ATACGTATAT TTTGTATTAC AGTGGGCACA CCCATGGTAC AGGAGAGTGG
1901 GCTCTAGCAG GTGGAGATAC ACTACGCCTT GACACACTTA TAGAATGGTG
1951 GAGAGAAAAG AATGGTTCCT TTTGTTCCCG GCTTATTATC GTATTAGACA
2001 GCGAAAATTC AACCCTTGG GTGAAAGAAG TGAGGAAAAT TAATGACCAG
2051 TATATTGCAG TGCAAGGAGC AGAGTTGATA AAAACAGTAG ATATTGAAGA
2101 AGCTGACCCG CCACAGCTAG GTGACTTTAC AAAAGACTGG GTAGAATATA
2151 ACTGCAACTC CTGTAATAAC ATCTGCTGGA CTGAAAAGGG ACGCACAGTG
2201 AAAGCAGTAT ATGGTGTGTC AAAACGGTGG AGTGACTACA CTCTGCATTT
2251 GCCAACGGGA AGCGATGTGG CCAAGCACTG GATGTTACAC TTTCTCGTA
2301 TTACATATCC CTTAGTGCAT TTGGCAAATT GGTATGCGG TCTGAACCTT
2351 TTTTGATCT GCAAACTTG TTTAGGTGTC TTGAAAAGAT TAAAAATGAG
2401 TTGGTTTCTT CTTACTGTGC TGGACACAGG ACAAGGCTTC AAACCTGTCA
2451 AATCTTAATT TGGACCCCAA AGCGGGATAT TAATAAGCAC TCATACTACC
2501 AATTATCACT AACTTGCCAT TTTTGTATG CTGTATTTT ATTTGTGGAA
2551 AATACCTTGC TACTTCTGTA GCTGCTCTCA CTTTGTCTTT TCTTAAGTAA
2601 TTATGGTATA TATAAGCGT TGGGAAAAAA CATTTTATAA TGAAAGTATG
2651 TAGGGAGTCA AATGCTTACT GTAAATGCAT AAGAGACGTT AAAAATAACA
2701 CTGCACTTTC AGGAATGTTT GCTTATGGTC CTGATTAGAA AGAACAGTT
```



```

2751 GTCTATGCTC TGCAATGGTC AATGATGAAT TACTAATGCC TTATTTTCTA
2801 GGCATATAAT AATAGTTTAG AGAATGTAGA CCAGATAAAT TTGTTTACTG
2851 TTTTAAGAAA ACTACCATG TACTTACAGA AGATTCTTTT TTCCAAACAG
2901 TAGGTTTCAT CCAAGACCAT TTGAAGAACT GCAAACCTCT TCTCTTAGAA
2951 AAGAAAGAGG GCAGCCTAAA ATAAACGCAA AATTGCTTA TACTCCATCA
3001 CATTAGATG TCTTGGTTGT GACTTATTAC CAGTGTGGCA GAGAACCCAA
3051 GTTACATTTT AGATCAAAAT ATTCTTTATG TAGGTATTGT TAAAAGGCTA
3101 GAGCCTACAA GTTGCTCTTC CATGCGTTGG TCAGGGGGCC CTGAAAACAC
3151 TGGTAATATT AAGAGTCTTT CTCAGGGTAA CTTAATGTTT TCTTAATGAA
3201 CAGTGTTCCT AGCTACAAAT TCTTCCAATA AATTGTCTTC CTTTTTGAAA
3251 AGTACTCTCA TAGAAGAAAT TTAGCAATT CTCTTGACT GACTCAGTCT
3301 ATTTTAAGTA TTCAGAAAAG ATTTTGATCC CCATTGAGT AATGCTCTGC
3351 CTTGAAAATT ATTTTCTGA TCCTTGTTAG TGATAACATT TTTTCTCTAC
3401 TGAAGGTCAG AGGATAGGAA ACAAGTATT CTCTTCTGGT ATACATGTAA
3451 TGTATTCTGT AAAAAAGTAT TCATATTGGC AATTTTAGTT AGGCATAATA
3501 TTGTGGTTGT AATTTTTTAA ACTTAGTGT TTGCTGATT AAAGCAGGCA
3551 CTGATCAGGG TATCTCCTAA GAGGTAATC ACTTCTTATT CCTTTCCAAT
3601 AATTATTACA TTCTAAATT TCATCTATGA GAAATAACAA ACAAGAAGGG
3651 AATAGAATTA AATTGGGGTA TAATCTAATC TTCATTGTTT AAATGGTTTG
3701 CCTTCTCACC ATTGAAGCCA TTTTCTTATA GCCTCAGAAA GAGGAAATAA
3751 TGCCTCCACC ATTTTCTACC TGGTGACTTG AAAATTGAAC TTTTAAGTTA
3801 GGAAGAAGTT AGAGTCAGGG AACTTGTATA CCACTATCTA TGCAGCATTG
3851 TTATAGTCTG ATTATTCTG TGTTTTGAAT ATGATTTTCC TAATGCTCTA
3901 AATAAAATTT TGTAAAAAT CAAAAAATAA AAAAAAATAA CTTATCGATA
3951 CCGTCGACCT CGATGATGTC GAC

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 365 bp to 2455 bp; peptide length: 697
 Category: putative protein
 Classification: unset

```

1 MCKSLRYCFS HCLYLAMTRL EEVNREVMNH SSVRYLGYLA RINLLVAICL
51 GLYVRWEKTA NSLILVIFIL GLFVLGIASI LYYYFSMEAA SLSSLNLWFG
101 FLLGLLCFLD NSSFKNDVKE ESTKYLLTSS IVLRILCSLV ERISGYVRHR
151 PTLTITVEFL ELVGFAIAST TMLVEKSLSV ILLVVALAML IIDLRMKSFL
201 AIPNLVIFAV LLEFSSLETP KNPIAFACFF ICLITDPFLD IYFSGLSVTE
251 RWKPFYLRGR ICRRLSVVFA GMIELTFEIL SAFKLDRTHL WYFVIPGFSI
301 FGIFRMICHI IFLLTLWGFH TKLNDCHKVY FTHRTDYNLS DRIMASKGMR
351 HFCLISEQLV FFSLLATAIL GAVSWQPTNG IFLSMFLIVL PLESMAGLGF
401 HELGNCLGGT SVGYAIVIPT NFCSPDQOPT LLPPEHVQEL NLRSTGMLNA
451 IQRFFAYHMI ETYGCYSTS GLSFDLHLSK LKAFLELRTV DGPRHDTYIL
501 YYSGHTHGTG EWALAGGDTL RLDTLIEWWR EKNGSFC SRL IIVLDSNST
551 PWVKEVRKIN DQYIAVQGAE LIKTVDIEEA DPPQLGDFTK DWVEYCNCS
601 NNICWTEKGR TVKAVYGVSK RWSDYTLHLP TGS DVAKHWM LHFPRITYPL
651 VHLANWLCGL NLFWICKTCF RCLKRLKMSW FLPTVLDTGQ GFKLVKS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphfbr2_2c1, frame 2

PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*, N = 1,
 Score = 96, P = 0.12

>PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*
 Length = 288

HSPs:

Score = 96 (14.4 bits), Expect = 1.3e-01, P = 1.2e-01
 Identities = 59/234 (25%), Positives = 116/234 (49%)

SEQ IIVLDSENSTPWVKEVRKINDQYIAVQGAEIKTVDIEEADPPQLGDFTKDWVEYNCNSC


```
SEG .....
PRD eeeeeccccccchhhhhccceeeccceeecccccccccccccccccccccc
MEM .....

SEQ NNICWTEKGRTVKAVYGVSKRWSYTLHLPTGSDVAKHWMHLHFPRITYPLVHLANWLCGL
SEG .....
PRD ceeeeccccceeeccccccccccccccccchhhhhccccccccchhhhhhhcc
MEM .....

SEQ NLFWICKTCFRCLKRLKMSWFLPTVLDTGQGFKLVS
SEG .....
PRD eeeeehhhhhhhhhhhhccceeeccccccccccccc
MEM .....
```

(No Prosite data available for DKFZphfbr2_2c1.2)

(No Pfam data available for DKFZphfbr2_2c1.2)

DKFZphfbr2_2c17

group: signal transduction

DKFZphfbr2_2c17.3 encodes a novel 446 amino acid protein with similarity to yeast YMR13lc and mammalian Retinoblastoma-binding protein RbAp46

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to YMR13lc and retinoblastoma-binding protein RbAp46

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2248 bp

Poly A stretch at pos. 2230, polyadenylation signal at pos. 2200

```
1 TGGGGAAGAT GGGCGCGCGC AAGGGTCGGC GTCGCACGTG TGAACCGGG
51 GAACCCATGG AAGCCGAGTC CGGCGACACA AGTTCCGAGG GCCCGGCCCA
101 GGTCTACCTG CCCGGCCGGG GGCCGCCGCT ACGCCAAGGG GAGGAGCTGG
151 TCATGGACGA GGAGGCCTAT GTGCTCTACC ACCGAGCGCA GACTGGCGCC
201 CCCTGTCTCA GCTTTGACAT AGTCCGGGAT CACCTGGGAG ACAACCGGAC
251 AGAGCTTCCT CTTACACTTT ACTTGTGTGC TGGGACCCAG GCTGAGAGCG
301 CCCAGAGCAA CAGACTGATG ATGCTTCGGA TGCACAATCT GCATGGGACA
351 AAGCCCCCAC CCTCAGAGGG CAGTGATGAA GAAGAAGAGG AGGAAGATGA
401 AGAGGATGAA GAAGAGCGGA AACCTCAGCT GGAGCTGGCC ATGGTGCCCC
451 ACTATGGTGG CATCAACCGA GTTCGGGTGT CATGGCTGGG TGAAGAGCCT
501 GTGGCTGGGG TGTGGTCAGA GAAGGGCCAG GTGGAGGTGT TTGCGCTGCG
551 GCGGCTTCTG CAGGTGGTGG AGGAGCCCCA GGCCCTGGCA GCCTTCTCTC
601 GGGATGAGCA GGCCCAAATG AAGCCCATCT TCTCCTTCGC TGGACACATG
651 GCGGAGGGCT TTGCCCTTGA CTGGTCCCCC CGGGTGACCG GTCGCCTGCT
701 GACCGGTGAC TGTCAAAGA ACATCCACCT CTGGACACCT ACGGACGGCG
751 GCTCCTGGCA CGTGGACCAG CGGCCATTCT TGGGCCACAC ACGCTCTGTG
801 GAGGACCTGC AGTGGTCACC GACTGAGAAC ACGGTGTTTG CCTCCTGCTC
851 AGCTGACGCC TCCATCCGCA TCTGGGACAT CCGGGCAGCC CCCAGCAAGG
901 CCTGCATGCT CACCACAGTC ACCGCCCATG ATGGGGACGT CAATGTCTATC
951 AGCTGGAGCC GCCGGGAGCC CTTCCTGCTC AGTGGCGGGG ATGATGGGGC
1001 CCTCAAGATC TGGGACCTTC GGCAGTTCAA GTCTGGTTCC CCAGTGGCCA
1051 CCTTCAAGCA GCACGTGGCC CCCGTGACCT CCGTCGAGTG GCACCCCCAG
1101 GACAGCGGGG TCTTTGCAGC CTCGGGTGCA GACCAACAGA TCACACAGTG
1151 GGACCTGGCA GTGGAGCGGG ACCCTGAGGC GGGCGACGTG GAGGCCGACC
1201 CCGGACTGGC CGACCTCCCG CAGCAGTGC TGTTCGTGCA CCAGGGCGAG
1251 ACCGAGCTGA AGGAGCTGCA CTGGCACCCG CAGTGCCCGG GGCTCCTGGT
1301 CAGCAGCGGG CTGTCAAGCT TCACCATCTT CCGCACCATC AGCGTCTGAG
1351 GCGTCCCACT GGCTCTGATC TTGCTTCTCT CTTGGAACCT GAAGTCTGAAT
1401 TGGGCTCCCC TGGGAAGGGT TCATTGAGT CTGTTGACTG AGACTGGCCG
1451 GCCTGTGGGC TGCCGTGATG GATTCTGTTT GACGTATTGT TCTCTAGAAG
1501 GCCTGGCTCT GATCCAGTGA CCCCTCTCAC CAAAGAACTC GGTTTAACCA
1551 GGGCTCTGTA AGACCACTCC CACCCAGAGA CTGTGTGGC CTGGTGTGGC
1601 CTGTGTGTCG GATTCTCTCC TGTGAGCTGT GACCCATTG ACCTGTGTCC
1651 CCAGAACCCA GTTTTGTGTT TGTGTTGTTG AGACGGAGTC TTGGTCTGTC
1701 GCCCAGGCTG GAGTGCAGTA GCACGATCTT GGCTCACTGC AACCTCCGCC
1751 TCCTGGGTTA AAGTGATTCT CTCAGCTCAG TCTCCCAGGT AGCTGGGATT
1801 ACAGGCATGT GCCACCACAC CCCGTTAATT TTTGTATTTT TAGTAGAGAC
1851 GGGGTTTCAC CATGTTGGCC AGGCTGGTCT CAAATTCTTG ATCTCAAGTG
1901 ATCTGTCCGC CCCGGCCTCC CAGAGTGCTG GGTGGGATT ACAGGCGTGA
1951 GCCACCGCGT CCGGCTCAGG ACCCAGTTT GGCTGCTGGT TCCAGCAGG
2001 GGAAGCTGGG GATATACAGT GGCTGCACCA AATTGGAGGT GTGGGTTCTT
2051 CCAACACAAT TTGCTTCTGC CCGTTGTCTT CTTGCCAGCT GGGTTTGGCC
2101 AGGATTCTCT CGTGTGGGGG CTACATGCGA CCTCTCCCC TCCTCCCTGA
2151 CTTTAGAGGC TGGTGTGTG TCGGGAGGAA GGTGAGGCT CCTGAGCAGC
2201 AATAAAGGAC CAGGAAGAGG CTGAGGTGG AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 9 bp to 1346 bp; peptide length: 446
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: WD_REPEATS (323-338)

```

1 MAARKGRRRT CETGEPMEAE SGGTSSEGPA QVYLPGRGPP LREGHEELVMD
51 EEAYVLYHRA QTGAPCLSFD IVRDHLGDNR TELPLTLYLC AGTQAESQAQ
101 NRLMLRMHN LHGTPPPSE GSDEEEEEED EDEEERKPKQ LELAMVPHYG
151 GINRVRVSWL GEEPVAGVWS EKGQVEVFAL RRLQVVEEP QALAAFLRDE
201 QAQMKPIFSF AGHMGEGFAL DWSRVGTGRLLTGDCQKNIH LWTPTDGGSW
251 HVDQRPFEVGH TRSVEDLQWS PTENTVFASC SADASIRIWD IRAAPSKACM
301 LTTVTAHDGD VNVISWSRRE PFLLSGGDDG ALKIWDLRQF KSGSPVATFK
351 QHVAPVTSVE WHPQDSGVFA ASGADHQITQ WDLAVERDPE AGDVEADPGL
401 ADLPQQLLFV HQGETELKEL HWHPQCPGLL VSTALSGFTI FRTISV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c17, frame 3

TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence., N = 1, Score = 910, P = 2.7e-91

PIR:S53061 hypothetical protein YMR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 691, P = 4.3e-68

PIR:I49367 retinoblastoma-binding protein mRbAp46 - mouse, N = 1, Score = 338, P = 1.1e-30

PIR:I39181 retinoblastoma-binding protein RbAp46 - human, N = 1, Score = 338, P = 1.1e-30

>TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence.
 Length = 469

HSPs:

Score = 910 (136.5 bits), Expect = 2.7e-91, P = 2.7e-91
 Identities = 195/442 (44%), Positives = 259/442 (58%)

```

Query: 18 EAESGDTSSSEGPAQVYLPGRGPPLREGEELVMDDEEAYVLYHRAQTGAPCLSFDIVRDHLG 77
      EA S + S P +V+ PG L +GEEL D AY H G PCLSFDI+ D LG
Sbjct: 18 EASSSEIPSI-PTRVWQPGVDT-LEDGEELQCDPSAYNSLHGPHVGVWPCLSFDILGDKLG 75

Query: 78 DNRTEPLTLYLCAGTQAESQAQSNRLMLRMHNLHGTP---PPSEGSDEEEEEDEED- 133
      NRTE P TLY+ AGTQAE A N + + ++ N+ G + P + G+ E+E+E+DE+D
Sbjct: 76 LNRTEFPHTLYMVAGTQAEKAAHNSIGLFKITNVSGKRRDVVPKTFGNGEDEDEDEDDDS 135

Query: 134 -----EEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKQVEVFALRRLQ 185
      E + P +++ V H+G +NR+R + W++ G V+V+ + L
Sbjct: 136 DSDDDDGDEASKTENIQVRRVAHGCNVRIRAMPQNSH-ICVSWADSGHVQVWDMSSHLN 194

Query: 186 VVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSRVGTGRLLTGDCQKNIHLWTPT 245
      + E + P+ +F+GH EG+A+DWSP GRLL+GDC+ IHLW P
Sbjct: 195 ALAESETEGKDGTSPLVNQAPLVNFSGHKDEGYAIDWSPATAGRLSGDCKSMIHLWEPA 254

Query: 246 DGGSWHVDQRPFEVGHTRSVEDLQWSPTEPTVFASCSADASIRIWDIRAAPSKACMLTTVT 305
      G SW VD PF GHT SVEDLQWSP E VFASCS D S+ +WDIR S A +
Sbjct: 255 SG-SWAVDPIPFAGHTASVEDLQWSPAENVFASCSVDGSAVWDIRLGKSPAL---SFK 310

Query: 306 AHDGDNVVISWSRREPFL-SSGDDGALKIWDLRQFKSGSPV-ATFKQHVAPVTSVEVHP 363
      AH+ DVNVISW+R +L SG DDG I DLR K G V A F+ H P+TS+EW
Sbjct: 311 AHNADNVVISWNRLASCMASGDDGTFSIRDLRLIKGGDAVVAHFEYHKHPITSIEWSA 370

```


Query: 364 QDSGVFAASGADHQITQWDLAVERDPE-----AGDVEADPGLADLPQQLLFVHOGETEL 417
 ++ A + D+Q+T WDL++E+D E A E DLP QLLFVHGG+ +L
 Sbjct: 371 HEASTLAVTSGDNQLTIWDLSEKDEEEAEFNAQTKELVNTFQDLPPQLLFVHGGQKDL 430

Query: 418 KELHWHHPQCPGLLVSTALSGFTIFRTISV 446
 KELHWH Q PG+++STA GF I ++
 Sbjct: 431 KELHWHNQIPGMIISTAGDGFNIMPYNI 459

Pedant information for DKFZphfbr2_2c17, frame 3

Report for DKFZphfbr2_2c17.3

[LENGTH] 446
 [MW] 49447.38
 [pI] 4.82
 [HOMOL] TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein";
 Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence. 1e-90
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YMR131c] 4e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YDL145c] 5e-09
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c]
 5e-09
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
 TAF90 - TFIID subunit] 6e-09
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YMR116c] 5e-08
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-08
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YLR429w] 3e-07
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]
 3e-06
 [FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 4e-06
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 4e-06
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c]
 2e-05
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w]
 2e-05
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 2e-05
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL011w] 3e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 5e-05
 [BLOCKS] BL00678
 [SCOP] d2trcb_2.51.3.1.1 Transducin (heterotrimeric G protein), gamm 5e-29
 [PIRKW] plasma 6e-07
 [PIRKW] duplication 4e-12
 [PIRKW] hormone 6e-07
 [PIRKW] transmembrane protein 1e-07
 [PIRKW] stomach 6e-07
 [PIRKW] actin binding 1e-07
 [PIRKW] leucine zipper 1e-07
 [PIRKW] signal transduction 2e-06
 [PIRKW] heterotrimer 2e-06
 [PIRKW] peripheral membrane protein 6e-07
 [PIRKW] GTP binding 2e-06
 [SUPFAM] WD repeat homology 1e-63
 [SUPFAM] yeast coatamer complex alpha chain 1e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 4e-07
 [SUPFAM] PRL1 protein 8e-09

[SUPFAM] MS11 protein 4e-12
 [SUPFAM] coatomer complex beta' chain 1e-09
 [PROSITE] WD_REPEATS 1
 [PFAM] WD domain, G-beta repeats
 [KW] All_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.14 %

```

SEQ  MAARKGRRRTCTETGEPMEAESGDTSSSEGPAQVYLPGRGPFLREGEELVMDEEAYVLYHRA
SEG  .....
lgotB .....

SEQ  QTGAPCLSFDIRVDHLDGNRTELPLTLYLCAQTQAESAQSNRLMLRMHNLHGTKPPPE
SEG  .....
lgotB .....

SEQ  GSDEEEEEDEEERKQLELAMVPHYGGINRVVSWLGEEPVAGVWSEKQVEVFAL
SEG  ..XXXXXXXXXXXXXXXXX.....
lgotB .....

SEQ  RRLQLVVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIH
SEG  .....
lgotB .....EECCCCCEEEEEETTT-TCEEEEEETTTTEE

SEQ  LWTPTDGGSWHVDQRPVGHTRSVEDLQWSPTEVTVFASCSADASIRIWDIRAAPSKACM
SEG  .....
lgotB EEEETTT---CEEEEECECCCCCEEEEEETTTCE-EEEEETTTTEEEEEETTT--TEEEE

SEQ  LTTVTAHDGVDNVISWSRREPFLLSGGDDGALKIWDLRQFKSGSPVATFKQHVAPVTSVE
SEG  .....
lgotB EECBTTBTCEEEEEETTTTTEEEEEETTTTEEEEE.....

SEQ  WHPQDSGVFAASGADHQITQWDLAVERDPEAGDVEADPGLADLPQQLLFVHQGETELKEL
SEG  .....
lgotB .....

SEQ  HWHPQCPLLVSTALSGFTIFRTISV
SEG  .....
lgotB .....

```

Prosites for DKF2phfbr2_2c17.3

PS00678 323->338 WD_REPEATS PDOC00574

Pfam for DKF2phfbr2_2c17.3

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVNCVaFSPDGGrWFIvSGSWDgTCRLWD*
 ++GH+ V ++ +SP + +++S S D ++R+WD

Query 257 EVGHTRSVEDLQWSPTEVTVFASCSADASIRIWD 290

24.88 304 336 1 34 dkf2phfbr2_2c17.3 similarity to YMR131c and retinoblastoma-binding protein RbAp46

Alignment to HMM consensus:
 Query *MrGHnnWVNCVaFSPDGGrWFIvSGSWDgTCRLWD*
 + H+++V+ +++S + ++SG++DG +++WD

dkf2phfbr2 304 VTAHDGVDNVISWSRREPFLLSGGDDGALKIWD 336

DKFZphfbr2_2c18

group: brain associated

DKFZphfbr2_2c18 encodes a novel 302 amino acid protein with weak similarity to cyclin-dependent kinase p130-PITSLRE.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to cyclin-dependent kinase p130-PITSLRE

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2835 bp

Poly A stretch at pos. 2817, polyadenylation signal at pos. 2796

```

1 TGGGGCGGAG GCGGAGGGAG TCCAGAGCCT TGAGCCCGGT GCTCCTCCCT
51 CGCGCAGCGG TGGCTCTGCG GCCGCTGGAG TAAACACTGC CTTTGTTCCT
101 TAGCGCCTCG TCTTTCGTCG CCCCGTGCCC TCACGCCGCC GGGCTCTGGC
151 CGGCCCGCCC TCGGTCTCTG AACCCCATTT CGGCTCGTGC CGTGCGGATG
201 CAGCTGCCCC GCCTGGGTTT GGGCATTGAG CGGGAGGAGG AGGAGGAGCG
251 GCGGCGCCTG GCGGCGATGC GATGGGGAAC TGCTGCTGGA CGCAGTGCTT
301 CGGACTGCTT CGCAAGGAAG CGGGGCGGCT GCAGCGAGTA GGCGGCGGCG
351 GAGGATCCAA GTATTTTAGA ACATGCTCAA GAGGTGAGCA CTTGACAATA
401 GAGTTTGAGA ATCTAGTAGA AAGTGATGAA GGGGAGAGCC CAGGAAGCAG
451 TCATAGGCCT CTTACTGAGG AAGAAATTGT TGACCTAAGA GAAAGGCATT
501 ATGATTCATG TGCCGAAAAA CAAAAGATC TTGATGAGAA AATTCAAAAA
551 GAGTTAGCCT TACAAGAAGA GAAGTTAAGA CTAGAAGAAG AAGCTTTATA
601 CGCTGCACAG CGTGAAGCAG CCAGGGCAGC AAAGCAGCGA AAGCTCTTGG
651 AGCAAGAAAG GCAGAGAATT GTGCAGCAAT ATCATCCTTC CAACAATGGA
701 GAATATCAAA GTTCAGGACC AGAAGATGAC TTCGAATCTT GTTTGAGAAA
751 TATGAAGTCA CAGTATGAAG TTTTTCGAAG TAGTAGACTC TCATCAGATG
801 CTACAGTTTT GACACCAAT ACAGAAAGCA GTTGTGATTT AATGACCAAA
851 ACTAAATCAA CTAGTGGAAA TGACGACAGC ACATCCTTAG ATCTAGAGTG
901 GGAAGATGAA GAAGGAATGA ATAGAATGCT TCCAATGAGA GAACGTTCCA
951 AAACAGAGGA AGACATTCTA CGGGCAGCAC TTAAGTATAG CAACAAGAAG
1001 ACTGGAAGTA ATCCTACATC AGCCTCTGAT GATTCCAATG GGCTGGAGTG
1051 GGAAATGAT TTTGTTAGTG CCGAAATGGA TGATAATGGA AATTCGGAGT
1101 ATTCTGGATT TGTAATCCTT GTATTAGAAC TGCTGATTC TGGCATAAGG
1151 CATCTGTACA CAGATCAACA GACTCGATAG GGTAAAATTG TGTGACCTTG
1201 TTTATCAGTT ATGACCAAAAT GTTAAAAACC AACTAGAATG TATAAGTGAT
1251 TGTGCTTAGC CTTTTTGTA GGGAGATGTG TAAGAAACCA TGCTGTAAAT
1301 GCTTATTTTA TTACAAAGGA GTAGGGATGA TAGGATCTGA ATTGATACAG
1351 AATTAAGTGC AATTTTCATCA TCTGCCTTCT GCTTTTCAAG ACCAATTTAA
1401 TGGTCTGTGC ATGTTACTGA TTAAATTAC TTTGTCTTGT CTTTATAGCA
1451 TTTCTGTTTA CTATGGTAGA TTTCCACTTT CAATTTTTAA AATTAATTTT
1501 ACTTTGAATG ATTTATGAAG CCTATTTTCT TGCTAACTA TGAAAAATAT
1551 AAGACTTTTT TGTTAATTCT CAGCCGATGT GAAGGAAGCA TGAGGAGGGA
1601 TCGTCAGACT CAGATTTAGA ATAGTGTTC CTTTCCAGC ATTATTATT
1651 TCTATGACTT CTTTGGATTT TATTATCTAA TAGTAAGTAC AGTTGATGTG
1701 GGTAGATGAC TCTAAGAAAT GCTGAAGTAT CGGCATTACA TGTGTTTATT
1751 TACATGTCCT AGTTTGATAA TGTTGATTCA ATCTGAACAA AAGATAATAT
1801 AAAAATAACC CTTTCAGAGT TGGACATTTC AAGTTGGTAA TAATAAAAAA
1851 TAATATTTAA GAAGATATAT ATATATATAT ATTTAGTTTT TTCCACTTCA
1901 TTTTACATGC CACTATATTG ACTTTAATTG ATATACAGTA TTAAGTTTTT
1951 AGGTGCCATT ATTTTAAAAA AATTCTATAT TTCCAATGAA CGATGTTAGA
2001 TTTTACACAG AACATATTCT CTGCATGATT TCAGAAAAGA AAATCTAAAA
2051 AGGTAAATAC GGTATTTCAA ATAAAAATCCT TTCTGGTATG AAAGGCTCCA
2101 TTGATTTTAT TAAGCCTTCC TTTACCTTGT AGTACAAGGT GCTTTAATGG
2151 GATAGAACCTA AGCATATCAA TATCTATAAC TGCATTTTGT GCTAGACAAT
2201 TACTGTTCTT TTCTCTAAAA TGTATATGTC AATTTACAAG GCCAGGGATA
2251 GAAAACACTC CATAATTGCT TTCCTTGATT TTGCTGAGGA TTTGGTATGA
2301 TTTTAGTAAG CAAACTGTTT TTTGGTTTTT CCTTAATGTT TTTAATTTTT
2351 TTTCTCTTGT CAACATGAC GGTGCATGTT CTATATAATA TAGGAAGGTC
2401 CAGATATAAA TAGTAACCTA AAGTTCTTGC TGTGCTTAAA AAAAAAATC
2451 ATGTGGCTCT TTCAATATTT GAACTGCTAA GCAATGACAT CTGTAGTTTT
2501 ATCTCCTTTT TTATGTCATA GAAATTAATA TGATACTTTA AATATGTAAA
2551 TATAATACAT TGGTAATGCT ATTATTATA TCTGTCTTAA CATAATTAA
2601 GTTGTAGCTG TGTCTTGAA ATATTTTAA GGTAACTTAT ATTCACATTG
2651 CCTGTGTTAA TGCTTTTAA GGTGTGATA CATCAGATGT ATATTTTTGG

```


2701 TTTGGCATAA GCTACGATTG TAATTTTCT TGGCTTTTGG TTCATAAAGA
 2751 ATTTTGTGAA GGAATGGTAA CAAATGGTAA TTTACAAATG GTTGTGAATA
 2801 AACACATTTT TACACTTAAA AAAAAAAAAA AAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 272 bp to 1177 bp; peptide length: 302
 Category: similarity to known protein

1 MGNCWTQCF GLLRKEAGRL QRVGGGGGSK YFRTCSRGEH LTIEFENLVE
 51 SDEGESPGSS HRPLTEEEIV DLRERHYDSI AEKQKDLDEK IQKELALQEE
 101 KLRLEEEALY AAQREAAARA KQRKLEQER QRIVQQYHPS NNGEYQSSGP
 151 EDDFESCLRN MKSQYEVFRS SRLSSDATVL TPNTSSCDL MTKTKSTSGN
 201 DDSTSLDLEW EDEEGMNRML PMRERSKTEE DILRAALKYS NKKTGSNPTS
 251 ASDDSNGLEW ENDFVSAEMD DNGNSEYSGF VNPVLELSDS GIRHSDTDQQ
 301 TR

BLASTP hits

Entry A55817 from database PIR:
 cyclin-dependent kinase p130-PITSLRE - mouse
 Length = 783
 Score = 123 (43.3 bits), Expect = 0.00013, P = 0.00013
 Identities = 53/197 (26%), Positives = 96/197 (48%)

Alert BLASTP hits for DKFZphfbr2_2c18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2c18, frame 2

Report for DKFZphfbr2_2c18.2

[LENGTH] 302
 [MW] 34281.39
 [pI] 4.73
 [PROSITE] MYRISTYL 5
 [PROSITE] CK2_PHOSPHO_SITE 12
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 3
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 13.58 %
 [KW] COILED_COIL 13.58 %

SEQ MGNCWTQCFGLLRKEAGRLQRVGGGGGSKYFRTCSRGEHLTIEFENLVESDEGESPGSS
 SEGxxxxx.....
 PRD cccccccchhhhhhhheeeccccccccceccccchhhhhhhcccccccccc
 COILS

SEQ HRPLTEEEIVDLRERHYDSIAEKQKDLDEKIQKELALQEEKLRLEEEALYAAQREAAARA
 SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
 PRD ccchhhhhhhhhhhccchhh
 COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ KQRKLEQERQRIVQQYHPSNNGEYQSSGPEDDFESCLRNKMSQYEVFRSSRLSSDATVL
 SEG xxxxxxxx.....
 PRD hhhhhhhhhhhhhccccccccccccccccchhhhhhhhhheeeccccccccc
 COILS CCCCCCCC.....


```

SEQ      TPNTESSCDLMTKTKSTSGNDDSTSLDLEWEDEEGMNRMLPMRERSKTEEDILRAALKYS
SEG      .....
PRD      cccccccccccccccccccccccccchhhhhhhccccccccchhhhhhhccccccccchhhhhhhhhhhc
COILS    .....

SEQ      NKKTGSNPTSASDDSNGLWENDFVSAEMDDNGNSEYSGFVNPVLELSDSGIRHSDTDQQ
SEG      .....
PRD      cccccccccccccccccccccccccceeeccccccccccccccccceeecccccccccccccccc
COILS    .....

SEQ      TR
SEG      ..
PRD      cc
COILS    ..

```

Prosites for DKFZphfbr2_2c18.2

PS00005	60->63	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	65->69	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	198->202	CK2_PHOSPHO_SITE	PDOC00006
PS00006	204->208	CK2_PHOSPHO_SITE	PDOC00006
PS00006	226->230	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	245->251	MYRISTYL	PDOC00008
PS00008	291->297	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2c18.2)

DKF2phfbr2_2d15

group: differentiation/development

DKF2phfbr2_2d15 encodes a novel 438 amino acid protein similarity to Mus musculus testis-specific Y-encoded-like protein (Tspyl1).

The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. The novel protein is a new member of the TSPY-SET-NAP1L1 family, which represents proteins closely related to TSPY. Therefore, the new protein seems to be involved in early spermatogenesis.

The new protein can find application in modulating early spermatogenesis.

strong similarity to testis-specific Y-encoded-like protein

complete cDNA, complete cds, EST hits
localisation: primer B does not match perfect

Sequenced by Qiagen

Locus: /map="729.2 cR from top of Chr6 linkage group"

Insert length: 3229 bp

Poly A stretch at pos. 3206, polyadenylation signal at pos. 3184

```

1 GGAGACTGTA GGGTGGGCGG TGCGAGCGGC GGTAGCTCC CAGTTCGGCC
51 TCTGAGGAAA ACGGGCGGTC GCCTGCGGTT GGTCCGACTG TTAGCAACAT
101 GAGCGGCGCTG GATGGGGTCA AGAGGACCAC TCCCTCCAA ACCCACAGCA
151 TCATTATTTC TGACCAAGTC CCGAGCGACC AGGACGCACA CCAGTACCTG
201 AGGCTCCGCG ACCAAAGCGA GCGGACACAG GTGATGGCGG AGCCGGGTGA
251 GGGAGGCTCG GAGACCGTCG CGCTCCCGCC TTCACCGCCT TCAGAGGAGG
301 GGGGCGTACC CCAGGATCCC GCGGGCCGTG GCGGTACTCC CCAGATCCGA
351 GTTGTGGGGG GTCGCGGTCA TGTGGCGATC AAAGCCGGGC AGGAAGAGGG
401 CCAGCCTCCC GCCGAAGGCC TGGCAGCCGC TTCTGTGGTG ATGGCAGCCG
451 ACCGCAGCCT GAAAAAGGGC GTTCAGGGTG GAGAGAAGGC CCTAGAAATC
501 TGTGGCGCCC AGAGATCCGC GTCTGAGCTG ACGGCGGGGG CGGAGGCTGA
551 GCGGAGGAGG GTGAAGACAG GAAAGTGCGC CACCGTCTCA GCAGCCGTGG
601 CTGAGAGGGA GAGCGCTGAG GTGGTGGTGA AGGAAGGCCT GGCGGAGAAG
651 GAGGTAATGG AGGAGCAGAT GGAGGTAGAG GAGCAGCCGC CAGAAGGTGA
701 AGAAATAGAA GTGGCGGAGG AGGATAGATT GGAGGAGGAG GCGAGGGAGG
751 AAGAAGGGCC CTGGCCTTTG CATGAGGCTC TCCGCATGGA CCCTCTGGAG
801 GCCATCCAGC TGGAACTGGA CACTGTGAAT GCTCAGGCCG ACAGGGCCTT
851 CCAACAGCTG GAGCACAGT TTGGGCGGAT GCGTCGACAC TACCTGGAGC
901 GGAGGAACCT CATCATTCAG AATATCCCGG GCTTCTGGAT GACTGCTTTT
951 CGAAACCACC CCCAGTTGTC CGCCATGATT AGGGGCCAAG ATGCAGAGAT
1001 GTTAAGGTAC ATAACCAATT TAGAGGTGAA GGAACCTAGA CACCCTAGAA
1051 CCGGTTGCAA GTTCAAGTTC TTCTTTAGAA GAAACCCTTA CTTCAGAAAC
1101 AAGCTGATTG TCAAGGAATA TGAGGTAAGA TCCTCCGGCC GAGTGGTGTG
1151 TCTTTCTACT CCAATTATAT GGCGCAGGGG GCATGAACCC CAGTCCCTCA
1201 TTCCGAGAAA CCAAGACCTC ATCTGCAGCT TCTTCACTTG GTTTTCAGAC
1251 CACAGCCTTC CAGAGTCCGA CAAAATTGCT GAGATTATTA AAGAGGATCT
1301 GTGGCCAAAT CCACTGCAAT ACTACCTGTT GCGTGAAGGA GTCCGTAGAG
1351 CCGGACGTCG CCCGCTAAGG GAGCCTGTAG AGATCCCCAG GCCCTTTGGG
1401 TTCCAGTCTG GTTAACATTT GCCCTTGGGA ATACTCCTGC ACAAGGTCTC
1451 CTACACCTTT CTGCTGGACC TGTGCTTGGG CATCAGCAAT GAGTATGCCT
1501 TCTATTGTGC TTTGTTTTTG CTGACTTTTC TGCACCCTGT TTCCTTTGGA
1551 TATTCAGTTC TCTCAACCTC AAGATTGAGA CGGTGGTGGG TATGCTTCTC
1601 CACTTCCATA TGACCTTCAT GCTGTTCTGG AATATCACAT GCTACGAGGT
1651 CATCCTTCAC ACTACTTGTA AGCCAAGCAA ATGATACTGT AGATTGTACT
1701 GCCTTTATCT GCACTGCTTG GACCTGTTT ATTCCCAGGG CCTCTGAAC
1751 GGTGTGCTGC ACTTGGATT CTAGCTTTGG GAGCCTGTTC CACCTACTCA
1801 GCTCTGCATT GAGCAGTATG GGCACATGCC CTGTGGACAG TTAAGTGACG
1851 TTAATGAAC TCAAGGAGAA AAGCAGTGAG CCACTGTGTT TGTGTGATTT
1901 ATGGTACTTC ATTGCTCTC CTTACCTCT AGTCACTTTC TATTGCTACC
1951 TGCCCTACAT TGGCTCCTGC CAAGTCCCT CTCTCTCCCT GTTTTCTCTT
2001 TTTTTTTTTT TTTTTTTTTT TTTTGGAGCG GAGGACGGAG TCTTGTCTCTG
2051 TCGCCCAAGT TGGAGTGAG TGGCGGATC TCGGCTCACT GCAACCTCCA
2101 CCTCCCGGGT TCAAGCGATT CTCCTGCCCT AGCCTCCCGA GTAGCTGGGA
2151 CTACAGGCGC GCGCCGCCAC GCCCGGCTAA TTTTATATT TTTAGTAGAG
2201 ACGGGGTTTC ACCATGCTGG CCAGGCTGGT CTCGAACCCC GACCTCGTGA
2251 TCCGCCCTCC TTAGCCTCCC AATCCTCTCT TAAAAAGTG ATAGCTCAGA
2301 AATATTGTGA AAAGCAAGGT TTTTATTTCA TTTTGGCTCT GTCATTTTCA
2351 GAGGCAAGAA AGTTGGCCTG TAAAATAGAG TGCTAGAGCT CTTACGCCCC
2401 TCCCCTTTCT CCCAACTTCC TACTTCCTAG CCCTTTTATC AACTCCTAGA
2451 ATAGTTAAAG AGAGACACAT CTAGATGGGA TGAAGGTGCT CTAAGCAGG

```



```

2501 AGAACTGAA CAAAAGGCTA GAGGCATGGG CCAGGTAAAA ATTGGGCCTA
2551 GAGTGAAGAC TGTGCTGCCG TTAAGAGCTT TCGAGGAAGG AGTACTTACT
2601 CCCCAATGAT GATGAATGGA GAAATACTTT TCAGGGAGAA TTGAAGGGGT
2651 TAAAGTGTTA AATATGTTGC CTAGACAAGG GTTCTTTAAA GAAAGACAGC
2701 GCAACTTTGA ATGCTTTTCTT ACTTGTTTTG TGACCTAATT TATGTGGAAG
2751 ATTTGTTATTT CATTAGGATT TAGTAAAAAT TTTTCTCTG ATTCTAAACT
2801 TATTGTGAAA ATTGAGCTGT ACAGATATTC TTTTGATTTC AATTGGGAAC
2851 ATTTGGAAGA ACAACAGTCT TACTTGCCTG TACAATATAG ACACATATGA
2901 ATAGTCATAA CAGTTTTTCAA CTTGTCTTGT TTTCTGTAA ACTATATTCC
2951 TAGAAACATA GTTTGAACAA CTTGGTCTTT GTTAGGCTTG TCAAATTGCC
3001 TTCAATGAAA AATAATCTAC AAAAGTATGG TTTAATTGAT TGTCTTACAT
3051 GATAATTTTC CCTGGCAACA ACTTAGTAAG TGATATATCT TTTTCTCTAA
3101 ATTGCTTAAA TACTGTGAAA TTGCTCTGAC AAATTGGAAG TGTACCATTG
3151 GCATATTGTT CTTCCTTTTT ATGCATGATG GTAAAAATAA AGCATGTTGT
3201 TCTGCTAAGA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry AF042181 from database EMBLNEW:
Homo sapiens testis-specific Y-encoded-like protein (TSPYL) mRNA,
partial cds.
Score = 3411, P = 6.9e-148, identities = 685/687

Entry HS938343 from database EMBL:
human STS WI-11947.
Score = 1195, P = 2.1e-46, identities = 273/299

Medline entries

98399864:
Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1 family

Peptide information for frame 3

ORF from 99 bp to 1412 bp; peptide length: 438
Category: strong similarity to known protein
Classification: Differentiation/Development

```

1 MSGLDGVKRT TPLQTHSIII SDQVPSDQDA HQYLRLRQDS EATQVMAEPG
51 EGGSETVALP PSPFSEEGGV PQDPAGRGGT PQIRVVGGRG HVAIKAGQEE
101 GQPPAEGGLA ASVVMADRS LKKGVGQGEK ALEICGAQRS ASELTAGAEA
151 EAEEVKTGKC ATVSAAVAER ESAEVVKEG LAEKEVMEEQ MEVEEQPPEG
201 EEIEVAAEDR LEEEAAREEG PWPLHEALRM DPLEAIQLEL DTVNAQADRA
251 FQOLEHKFGR MRRHYLERRN YIIQNIPIGW MTAFRNHPQL SAMIRGQDAE
301 MLRYITNLEV KELRHPRTGC KFKFFERRNP YFRNKLVKE YEVRSRGRV
351 SLSTPIIWRG GHEPQSFIIR NQDLICSFET WFDHSLPES DKIAEIIKED
401 LWPNPQLQYL LREGVRRARR RPLREPVEIP RPFQFQSG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d15, frame 3

TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds., N = 1, Score = 1202, P = 3.1e-122

TREMBL:AB018264_1 gene: "KIAA0721"; product: "KIAA0721 protein"; Homo sapiens mRNA for KIAA0721 protein, partial cds., N = 1, Score = 798, P = 2e-79

TREMBL:AB015345_1 gene: "HRIHFB2216"; Homo sapiens HRIHFB2216 mRNA, partial cds., N = 1, Score = 570, P = 2.9e-55

>TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds.
Length = 379

HSPs:

Score = 1202 (180.3 bits), Expect = 3.1e-122, P = 3.1e-122
Identities = 258/377 (68%), Positives = 283/377 (75%)

```

Query:   62 SPPSEEGGVQDPAGR-----GGTPQIRVVGGRGHVAIKAGQEE--GQP-P--AEGGLAA 110
          SP +EG   D G           GTP R + G           G+   G P P   EGL
Sbjct:   3 SPERDEGTPVPDSRGHCDADTVSGTPDRRPLLGEEKAVTGEGRAGIVGSPAPRDVEGLVP 62

Query:   111 ASVVMAADRSLLK-GVQGGKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAE 169
          V AA +       V+G   A+ +       ++ T GAE++A +VKT +   TV+AA
Sbjct:   63 QIRVAAARQGESPPSVRGPAAAVFVTPKYVEKAQETRGAESQARDVKT-EPGTVAATA-- 119

Query:   170 RESAEVVVKEGLAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWLHEALR 229
          E +EV           EE MEVE Q P GEE+E+ E           EA EE GPW L   LR
Sbjct:   120 -EKSEVATPGS-----EEVMEVE-QKPAGEEMEMLEASGGVREAPPEAGPWHLGIDLR 170

Query:   230 MDPLEAIQLELDTVNAQADRAFOQLEHKFGRMRHYLERRNYIIQNIPGFWMTAFRNHPQ 289
          +PLEAIQLELDTVNAQADRAFO LE KFGMRMRHYLERRNYIIQNIPGFWMTAFRNHPQ
Sbjct:   171 RNPLEAIQLELDTVNAQADRAFOHLEQKFGMRMRHYLERRNYIIQNIPGFWMTAFRNHPQ 230

Query:   290 LSAMIRGQDAEMLRITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLIKEYEVRSSGRV 349
          LSAMIRG+DAEMLRIT+LEVKELRHP+TGCKFKFFFRNPNPYFRNKLIKEYEVRSSGRV
Sbjct:   231 LSAMIRGRDAEMLRITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLIKEYEVRSSGRV 290

Query:   350 VSLSTPIIWRRGHEPQSFIRNQLDLCFFTFWSDHSLPESDKIAEIIKEDLWPNPLQYY 409
          VSLSTPIIWRRGHEPQSFIRNQLDLCFFTFWSDHSLPESD+IAEIIKEDLWPNPLQYY
Sbjct:   291 VSLSTPIIWRRGHEPQSFIRNQLDLCFFTFWSDHSLPESDRIAEIIKEDLWPNPLQYY 350

Query:   410 LLREGVRRARRRPLREPVEIPRPFQSG 438
          L REG+RR RRRP+REPVEIPRPFQSG
Sbjct:   351 LCREGIRRRRRRPIREPVEIPRPFQSG 379

```

Pedant information for DKFZphfbr2_2d15, frame 3

Report for DKFZphfbr2_2d15.3

```

[LENGTH]      438
[MW]           49307.65
[pI]           5.36
[HOMOL]        TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like
protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds. 1e-
107
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR048c]
1e-07
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR048c] 1e-07
[BLOCKS]       BL00376F
[PIRKW]        nucleus 6e-39
[PIRKW]        DNA binding 3e-06
[PIRKW]        phosphoprotein 6e-39
[PIRKW]        alternative splicing 6e-39
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY 22.83 %

```

```

SEQ  MSGLDGVKRTTFLQTHSIIISDQVPSDQDAHQYLRLRDQSEATQVMAEPGEGGSETVALP
SEG  .....X
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  PSPPSEEGGVQDPAGRGGTQIRVVGGRGHVAIKAGQEEGQPPAEGLAASVVMAADRS
SEG  xxxxxxxx.....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LKKGVGQGGKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAERESAEVVVKEG
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWLHEALRMDPLEAIQLEL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.
PRD  hhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  DTVNAQADRAFOQLEHKFGRMRHYLERRNYIIQNIPGFWMTAFRNHPQLSAMIRGQDAE
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  MLRYITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLIKEYEVRSSGRVVSLSTPIIWR

```



```
SEG .....
PRD hhhhhhhhhhhhhccccceeeeeccccccchhhhhccccccccccccceeeccc

SEQ GHEPQSFIRRNQDLICSFFTWFSDHSLPESDKIAEIIKEDLWPNPLQYYLLREGVRRARR
SEG .....xxxxxxxxxxxx
PRD ccccchhhhhccccceeeeeccccccchhhhhhhhhccccceeeccccchhhh

SEQ RPLREPVEIPRPFQSG
SEG xxxxxxxx.....
PRD hcccccccccccccccc
```

(No Prosite data available for DKFZphfbr2_2d15.3)

(No Pfam data available for DKFZphfbr2_2d15.3)

DKFZphfbr2_2d17

group: transmembrane proteins

DKFZphfbr2_2d17 encodes a novel 292 amino acid protein with similarity to a C.elegans hypothetical protein.

One transmembrane region is predicted for the protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans hypothetical protein

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 1009 bp

Poly A stretch at pos. 990, polyadenylation signal at pos. 969

```

1 TGGGCCTGTG GCTGGGGGCA GAGCTCAGAC TGTCTTCTGA AGATTGATGT
51 CTATTTCCTT GAGCTCTTTA ATTTTGTTGC CAATTTGGAT AAACATGGCA
101 CAAATCCAGC AGGGAGGTCC AGATGAAAAA GAAAAGACTA CCGCACTGAA
151 AGATTTATTA TCTAGGATAG ATTTGGATGA ACTAATGAAA AAAGATGAAC
201 CGCCTCTTGA TTTTCCTGAT ACCCTGGAAG GATTGGAATA TGCTTTTAAT
251 GAAAAGGGAC AGTTAAGACA CATAAAAACT GGGGAACCAT TTGTTTTTAA
301 CTACCGGGAA GATTTACACA GATGGAACCA GAAAAGATAC GAGGCTCTAG
351 GAGAGATCAT CACGAAGTAT GTATATGAGC TCCTGGAAAA GGATTGTAAT
401 TTGAAAAAAG TATCTATTCC AGTAGATGCC ACTGAGAGTG AACCAAAGAG
451 TTTTATCTTT ATGAGTGAGG ATGCTTTGAC AAATCCACAG AAAGTATGGG
501 TTTTAAATCA TGGTAGTGGT GTTGTGAGG CAGGGCAGTG GGCTAGAAGA
551 CTTATTATAA ATGAAGATCT GGACAGTGGC ACACAGATAC CGTTTATTAA
601 AAGAGCTGTG GCTGAAGGAT ATGGAGTAAT AGTACTAAAT CCCAATGAAA
651 ACTATATTGA AGTAGAAAAG CCGAAGATAC ACGTACAGTC ATCATCTGAT
701 AGTTCAGATG AACCAAGCAG AAAACGGGAA AGAAAAGATA AAGTTCTTAA
751 AGTAACAAAG AAGCGACGTG ATTTCTATGA GAAGTATCGT AACCCCAAAA
801 GAGAAAAAGA AATGATGCAA TTGTATATCA GAGTGAGTGA GATCACTACT
851 TTCCTTTACT ATTTTCTTTA CCTTGTATAT ATTTTATTAT ATGTAGATTG
901 TTTTGTTTT CTCAAGAAT ATTAATTTCT TTATTTGTCA TCATTATTTT
951 CCCATGGTCG TCTACTTGGG TTAATGGGT TTTTAAATTC AAAAAAATAA
1001 AAAAAAATAA

```

BLAST Results

Entry I89937 from database EMBL:
Sequence 11 from patent US 5723315.
Score = 1083, P = 2.2e-42, identities = 223/231

Entry I89938 from database EMBL:
Sequence 12 from patent US 5723315.
Score = 875, P = 7.4e-33, identities = 175/175

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 922 bp; peptide length: 292
Category: similarity to unknown protein
Classification: unset

1 MSISLSSLIL LPIWINMAQI QGGPDEKEK TTALKDLLSR IDLDELMKKD


```

51 EPPLDFPDTL EGFEYAFNEK QLRHIKTGE PFVFNRYREDL HRWNQKRYEA
101 LGEIITKYVY ELLEKDCNLK KVSIPVDATE SEPKSFIFMS EDALTNPQKL
151 MVLINGSGVV RAGQWARRLI INEDLDSGTQ IPFIKRAVAE GYGIVLNP
201 ENYIEVEKPK IHVQSSSDSS DEPAEKREK DKVSKVTKKR RDFYEKYRNP
251 QREKEMMLY IRVSEITTFLL YFFLYLVYIL LYVDCFVFLQ EY

```

BLASTP hits

Entry S67436 from database PIR:
 hypothetical protein - fission yeast (*Schizosaccharomyces pombe*)
 Length = 266
 Score = 112 (39.4 bits), Expect = 0.00037, P = 0.00037
 Identities = 33/147 (22%), Positives = 69/147 (46%)

Entry CEY75B8A.12 from database TREMBLNEW:
 gene: "Y75B8A.31": *Caenorhabditis elegans* cosmid Y75B8A
 Score = 327, P = 1.5e-29, identities = 72/140, positives = 93/140

Alert BLASTP hits for DKFZphfbr2_2d17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d17, frame 2

Report for DKFZphfbr2_2d17.2

```

[LENGTH]      292
[MW]           34260.50
[pI]           5.50
[HOMOL]        TREMBLNEW:AF064782_1 product: "unknown"; Mus musculus clone pEN87 unknown mRNA,
partial cds. 1e-119
[KW]           SIGNAL PEPTIDE 19
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 10.96 %

```

```

SEQ  MSISLSSLLLPWINMAQIQGGPDEKEKTTALKDLLSRIDLDELMMKKDEPPLDFPDTL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  EGFEYAFNEKQLRHIKTGEPFVFNRYREDLHRWNQKRYEALGEIITKYVYELLEKDCNLK
SEG  .....
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  KVSIPVDATESEPKSFIFMSDALTNPQKLMVLINGSGVVVRAGQWARRLIINEDLDSGTQ
SEG  .....
PRD  eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeecccccccccccccccccccc
MEM  .....

SEQ  IPFIKRAVAEGYGIVLNPENYIEVEKPKIHVQSSSDSSDEPAEKREKDKVSKVTKKR
SEG  .....
PRD  chhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RDFYEKYRNPQREKEMMLYIRVSEITTFLLYFFLYLVYILLYVDCFVFLQ EY
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

(No Prosite data available for DKFZphfbr2_2d17.2)

(No Pfam data available for DKFZphfbr2_2d17.2)

DKFZphfbr2_2d20

group: brain derived

DKFZphfbr2_2d20 encodes a novel 197 amino acid protein with similarity to *Synechocystis* sp. P74594 hypothetical 32.8 kD protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *Synechocystis* sp. (PCC 6803)

complete cDNA, complete cds, EST hits
potential start at bp 67 matches kozak consensus A NCatgG

Sequenced by Qiagen

Locus: unknown

Insert length: 1787 bp

Poly A stretch at pos. 1768, polyadenylation signal at pos. 1743

```
1 TGGGGCGGCC GCGGCGGGAA CATGGAGGAG CTGCTGAGGC GCGAGCTGGG
51 CTGCAGCTCT GTCAAGGCCA CGGGCCACTC GGGGGGCGGG TGCATCAGCC
101 AGGGCCGGAG CTACGACACG GATCAAGGAC GAGTGTTCGT GAAAGTGAAC
151 CCCAAGGCGG AGGCCAGAAG AATGTTTGAA GGTGAGATGG CAAGTTTAAC
201 TGCCATCCTG AAAACAAACA CGGTGAAAGT GCCCAAGCCC ATCAAGGTTG
251 TGGATGCCCC AGGCGGCGGG AGCGTGCTGG TGATGGAGCA CATGGACATG
301 AGGCATCTGA GCAGTCATGC TGCAAAGCTT GGAGCCCAGC TGGCCGATTT
351 ACACCTTGAT AACAAGAAGC TTGGAGAGAT GCGCCTGAAG GAGGCGGGCA
401 CAGTGTGGAG AGGAGGTGGG CAGGAGGAAC GGCCCTTTGT GGCCCGGTTT
451 GGATTTGACG TGGTGACGTG CTGTGGATAC CTCGCCAGG TGAATGACTG
501 GCAGGAGGAG TGGGTCGTGT TCTATGCCCG GCAGCGCATT CAGCCCCAGA
551 TGGACATGGT GGAGAAGGAG TCTGGGGACA GGGAGGCCCT CCAGCTTTGG
601 TCTGCTCTGC AGTAAAAGAT CCCTGACCTG TTCCGTGACC TGGAGATCAT
651 CCCAGCCTTA CTCCACGGGG ACCTCTGGGG TGGAAACGTA GCAGAGGATT
701 CCTCTGGGCC GGTGATTTT GACCCAGCTT CTTTCTACGG CCACTCGGAA
751 TATGAGCTGG CAATAGCTGG CATGTTTGGG GGCTTTAGCA GCTCCTTTTA
801 CTCGCCCTAC CACGGCAAAA TCCCAAGGC CCCAGGATTC GAGAAGCGCC
851 TTCAGTTGTA TCAGCTCTTT CACTACTTGA ACCACTGGAA TCATTTTGGA
901 TCGGGGTACA GAGGATCCTC CCTGAACATC ATGAGGAATC TGGTCAAGTG
951 AGCGGGCCTT ACTCTGGAAG GAGGTCTCAG AGGTTTCTCC ACAGTCCTCT
1001 TCTGGGCAAA TTCTTGTTTC TTCACATGCC GGA CTAGCTT AAGACCAATG
1051 CAGTAGCTTA TTTCCAAGCC TTGCAAAAGTA TATAATATCT AAGAGGAAAG
1101 GTTTTGTGTC CCCAGCGTTG TCCACTTTGT GGGGCTTTGT AGGTAGACGG
1151 AGCCACACTA CAGGCAGGGT ATGAGCAGAG GGAATGATGG AGTGTGGGCG
1201 ACTCTGAGCC TCACTGCTGC TGCAAGGTGG GGAAACTGTA AGTGAACCCC
1251 TGTGGGTGCG GGGGAGGGTA TCCGGTGCGC AGGGAGGTGG CCAGCGCCCC
1301 CGGGCACTGC TGCTCATAGG TACCTTTCCG CTGCCTCCTC CCGTCTCTCC
1351 TGTGCAGGAA TGTCTCTGAG CTGTTACAGT TGATGCTTCT TGGTTGGCAA
1401 GACTTGGGTG TAGACATGAA ACCACCTTAC TAAAAGCGTC TTAATAATGAC
1451 CAATTCCAGA ATCAAGCGTA TTCCGTTTTC CTCCTGCATG ATCCCTGGGC
1501 CCTCCCGCAG GCTGAGCAAG TCTGTAAACT GATTCTGGGA GAAACCAAGC
1551 TGCTGGCCGT AGGATGTCCT TGGGTACATC CAGGAGTCTT CATTGCTTCT
1601 GTTATTACCC CGTCTCCTCT GCCATTTTCT ACAGCTTGCT GAGTTGTCAT
1651 TCCCTTGCAA CATTAAAATA CATGCTGAAC TCATATTTT CTTCCTTCA
1701 CTGTTGTAGT AAAGAGACAT ATTTTCATGAA TGGCATTGAT GCTAATAAAC
1751 CCTTGGCCCA AAAATTGAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 612 bp; peptide length: 197
 Category: similarity to unknown protein
 Prosite motifs: LEUCINE_ZIPPER (117-139)

```

1 MEELLRRELG CSSVRATGHS GGGCISQGRS YDTDQGRVFV KVNPKAEARR
51 MFEGEMASLT AILKTNNTVKV PKPIKVLDP GGGSVLVMEM MDMRHLSSHA
101 AKLGAQLADL HLDNKKLGEM RLKEAGTVWR GGGQEERPFV ARFGFDVVT
151 CGYLPQVNDW QEDWVVFYAR QRIQPQMDMV EKESGDREAL QLWSALQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d20, frame 1

Report for DKFZphfbr2_2d20.1

```

[LENGTH]      197
[MW]           21963.25
[PI]           6.96
[HOMOL]        PIR:S76790 hypothetical protein - Synechocystis sp. (strain PCC 6803) 9e-12

[SUPFAM]       hypothetical protein b1725 1e-06
[PROSITE]      LEUCINE_ZIPPER 1
[PROSITE]      MYRISTYL 2
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 2
[KW]           Alpha_Beta

```

```

SEQ  MEELLRRELGCSSVRATGHSGGGCISQGRSYDTDQGRVFVKVNPKAEARRMFEGEMASLT
PRD  ccchhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhh

SEQ  AILKTNNTVKVPKPIKVLDPGGGSVLVMEHMDMRHLSSHA AKLGAQLADLHLDNKKLGEM
PRD  hhhhhhheeeccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhh

SEQ  RLKEAGTVWRGGQEERPFVARFGFDVVTCCGYLPQVNDWQEDWVVFYARQRIQPQMDMV
PRD  hhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhh

SEQ  EKESGDREALQLWSALQ
PRD  hhhccchhhhhhhhhhhcc

```

Prosite for DKFZphfbr2_2d20.1

PS00002	20->24	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	13->16	PKC_PHOSPHO_SITE	PDOC00005
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00029	96->118	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_2d20.1)

DKF2phfbr2_2g18

group: brain derived

DKF2phfbr2_2g18 encodes a novel 229 amino acid protein with partial similarity to the humane dJ30M3.2 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

J30M3.2 extension of genmodel

complete cDNA, complete cds, EST hits
(mouse ESTs with >90% Identities)

Sequenced by Qiagen

Locus: /map="6p22.1-22"

Insert length: 2444 bp

Poly A stretch at pos. 2425, no polyadenylation signal found

```
1 TGGTCGAGGG TCGACGGTAT CGATAAGTTT TTTTTTTTTT TTTTTTTTTT
51 TGGAAGCAAA GGATCACACT TCCCCTCCC TGTTCTTAA TCCCTTTTCT
101 AAAAAGGGGG GAAATCCGG ATGGATTTA GGGATTGGTC TGGTGTGAGC
151 TGTGTCTTAT TGCACACCTA AATCCTGATT ATAGGCTTTT CATTTCCTCG
201 CAAAGCCTTT ATTTGGCAG TTAAGCCAAA TGTGTTTCC AGAAAGTTAG
251 TTATTTCTC CTCTTCTTT CCTTCTTTC CTCCTTTT CCCGCTGAC
301 CCCAAACGTT ATTGTCCAAA CATGACTGGA CAGCAGCTTT TGTTCCTTGA
351 CCCTGTAATA TGACAGTCTG CTAATATTGA CAGAAGGTGC AGTTTTTGGG
401 TTATAGTCGT GATTTCGCT AATCAATCAT ATTAGCAGGA AAAAAAATGA
451 CTTGTTCTG TTGTACTTGA GTCTTAAGAA AAAGTGCCCA TAGTTAGTG
501 ACAATTTCCA AAGGCTTTAG TACCACCTGT ATTTCAAAAT GGGGGACCCA
551 AACTCCCGGA AGAAACAAGC TCTGAACAGA CTACGTGCTC AGCTTAGAAA
601 GAAAAAGAA TCTCTAGCTG ACCAGTTTGA CTTCAAGATG TATATGCCT
651 TTGTATTCAA GGAGAAGAAG AAAAAGTCAG CACTTTTGA AGTGTCTGAG
701 GTTATACCAG TCATGACAAA TAATTATGAA GAAAAATATCC TGAAAGGTGT
751 GCGAGATTCC AGCTATTCTT TGGAAAGTTC CCTAGAGCTT TTACAGAAGG
801 ATGTGGTACA GCTCCATGCT CCTCGATATC AGTCTATGAG AAGGGATGTA
851 ATTGGCTGTA CTCAGGAGAT GGATTTCATT CTTTGGCCTC GGAATGATAT
901 TGAAAAAATC GTCTGTCTCC TGTTTTCTAG GTGGAAGAA TCTGATGAGC
951 CTTTAGGCC TGTTCAAGCC AAATTGAGT TTCATCATGG TGAATATGAA
1001 AAACAGTTTC TGCACTGACT GAGCCGCAAG GACAAGACTG GAATCGTTGT
1051 CAACAATCCT AACCACTCAG TGTTCCTCTT CATTGACAGA CAGCACTTGC
1101 AGACTCCAAA AAACAAAGCT ACAATCTTCA AGTTATGCAG CATCTGCCTC
1151 TACCTGCCAC AGGAACAGCT CACCACTGG GCAGTTGGCA CCATAGAGGA
1201 TCACCTCCGT CCTTATATGC CAGAGTAGAG TACTGACCAG CAAAATGGAG
1251 AAGATCAGAG AATGCAGCAG CAGTTTTTTT TCTTGTTTTC TTACCACTTT
1301 ATTCTTTCAG AGTTAAAGA AAATGGACTC ATGCACAGAA CACTATGCAT
1351 TTTGAACTT GTTCATCTCT GATTTTTTTA AATCATTTT ATCTCAGAAC
1401 TTAACAAAAA ATTAGATGTC GTGCACGGAC TGTGTGAAAG AAGATGCTTT
1451 GCATATTGTC TGCCTGTCAT CAGTATCTTA CTAAAAATGT GAAATGAAAG
1501 GACTATTGTA CACTGAAATG CTAAATGTA TCTGAAAGCA CAAGGTGATA
1551 CTCATTTTTA TGGTCTTCCC ATTTGTGCTG GTTTTTGCCT CTTTGACATC
1601 TGTATCATGT ATTTAGAGGG TGAGAAGTGA ATGTAACAGG TATAATAAC
1651 ATTTTAAAAA ACAATAACTT TGCTATAATC ACAGTTGTTT CAGAGCACTG
1701 TCAGATACAT TCTAATGACC AGAAGTGGTT TAAAAAAGA AAATACAACC
1751 ATGGGAAAGA AATCTTAAAT GAAAAACGCA TCTCATTTGA GGCATTTTTG
1801 CCTCATATTT TACTGGGCCA TGTTTGTTC CTGGTACTCA TGTATTTTTT
1851 TTTTTTCCAG ATCTCTTCC CCAAGTTGCT ATTGTAAGAG TATTCTGCTG
1901 CGTGTGGATG CAGTTATACA CATTAAAGCA GATCTGGAGT CTGAAGTAGC
1951 TATAAAGCAG CTATAAACA GAAATACATG CATAGCTGCA GAAACCATGA
2001 TAGGTAGAGG ACTTTCTTTT TGGTTTGTG TTTTGTGTT TTTTGTGTT
2051 TTTGGTTTTA CAGAGAAGAG ATTTTATTA CAAAGAAAAA AATCCAGTG
2101 AATGTGTCAG AAATGCTGGT TTTTACACCA TCCTAAAGAA AAACCTTACA
2151 AGGGTGTGTT GGAGTAGAAA AAAGTTTATA AAGTTGGAAT CTAAATTTGT
2201 AAAATTAAAC ATTGAGTGTC AAAGTTCTAA AAGCAGAACT CATTTCTGTC
2251 AATGAACATA AGGAAGACT ACTGTATAGG TTTTTTTTTT TCTCCTTTTA
2301 AATGAAGAAA AGCTTGCTT AAGGGTTGCA TACTTTTATT GGAGTAAATC
2351 TGAATGATCC TACTCCTTTG GAGTAAGACT AGTGCTTACC AGTTTCCAAT
2401 TGTATTAGC TTCTGTTGGA ATTTGAAAAA AAAAAAATAA AAAA
```

BLAST Results

Entry HS338352 from database EMBL:

human STS EST171398.

Score = 1747, P = 3.0e-74, identities = 359/365

Entry HS447255 from database EMBL:

human STS SHGC-10143.

Score = 1717, P = 6.5e-73, identities = 365/383

Entry HS30M3 from database EMBLNEW:

Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.

Score = 6646, P = 0.0e+00, identities = 1344/1355

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 539 bp to 1225 bp; peptide length: 229

Category: putative protein

```
1 MGDPNRSRKKQ ALNRLRAQLR KKKESLADQF DFKMYIAFVF KEKKKKSALF
51 EVSEVIPVMT NNYEENILKG VRDSSYSLES SLELLQKDVV QLHAPRYQSM
101 RRDVIGCTQE MDFILWPRND IEKIVCLLFS RWKESDEPFR PVQAKFEFHH
151 GDYEKQFLHV LSRKDKTGIV VNNPNQSVFL FIDRQHLQTP KNKATIFKLC
201 SICLYLPQEQ LTHWAVGTIE DHLRPYMPE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2g18, frame 2

TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands., N = 1, Score = 470, P = 1.1e-44

>TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands. Length = 86

HSPs:

Score = 470 (70.5 bits), Expect = 1.1e-44, P = 1.1e-44
Identities = 86/86 (100%), Positives = 86/86 (100%)

```
Query: 144 AKFEFHG DY EKQFLHVL SRKDKTGIV VNNPNQSVFL FIDRQHLQTPKNKATIFKLC SIC 203
      AKFEFHG DY EKQFLHVL SRKDKTGIV VNNPNQSVFL FIDRQHLQTPKNKATIFKLC SIC
Sbjct: 1 AKFEFHG DY EKQFLHVL SRKDKTGIV VNNPNQSVFL FIDRQHLQTPKNKATIFKLC SIC 60
```

```
Query: 204 LYLPQEQ LTHWAVGTIEDHLRPYMPE 229
      LYLPQEQ LTHWAVGTIEDHLRPYMPE
Sbjct: 61 LYLPQEQ LTHWAVGTIEDHLRPYMPE 86
```

Pedant information for DKF2phfbr2_2g18, frame 2

Report for DKF2phfbr2_2g18.2


```

SEQ      MGDPNRKRKQALNRLRAQLRKKKESLADQDFDKMYIAFVFEKKKKSALFEVSEVIPVMT
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeec

SEQ      NNYEENILKGVDRSSYSLESSLELLQKDVVQLHAPRYQSMRRDVGCTQEMDFILWPRND
SEG      ..... xxxxxxxxxxxxxx
PRD      cchhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhccccccccceeeccccceeeccch

SEQ      IEKIVCLLFSRWKESDEFFRPVQAKFEFHGDEYEQFLHVLRSRKDTGIVVNNPNQSVFL
SEG      .....
PRD      hhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhccccceeeccccceeee

SEQ      FIDRQHLQTPKNKATIFKLCSICLYLPQEQLTHWAVGTIEDHLRPMPE
SEG      .....
PRD      eeecccccccccccccccccccccccccccccccccccccccccccccccccc

```

PS000001	175->179	ASN_GLYCOSYLATION	PDOC000001
PS000004	22->26	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	44->48	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	6->9	PKC_PHOSPHO_SITE	PDOC000005
PS000005	99->102	PKC_PHOSPHO_SITE	PDOC000005
PS000005	162->165	PKC_PHOSPHO_SITE	PDOC000005
PS000005	189->192	PKC_PHOSPHO_SITE	PDOC000005
PS000006	25->29	CK2_PHOSPHO_SITE	PDOC000006
PS000006	80->84	CK2_PHOSPHO_SITE	PDOC000006
PS000006	162->166	CK2_PHOSPHO_SITE	PDOC000006
PS000006	218->222	CK2_PHOSPHO_SITE	PDOC000006
PS000007	69->77	TYR_PHOSPHO_SITE	PDOC000007
PS000008	70->76	MYRISTYL	PDOC000008
PS000008	168->174	MYRISTYL	PDOC000008

227

DKFZphfbr2_2h1

group: brain derived

DKFZphfbr2_2h1 encodes a novel 180 amino acid protein with weak similarity to C.elegans D2007.4 protein

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C.elegans D2007.4 protein

CpG island in 5' region, complete cDNA

Sequenced by Qiagen

Locus: unknown

Insert length: 957 bp

Poly A stretch at pos. 939, polyadenylation signal at pos. 916

```

1 GGGGGTCCCT GACTTTATAT GGCTGCTCCT GCGAGCGAC TGAGTCGTCC
51 GTGAGGAAAA AGAGGCGAGG CTTTCCGAG ATCGTCTCAG CGATGGCGCT
101 TCGGTCGCGG TTTTGGGGGT TGTTCTCGGT TTGCAGGAAC CCTGGGTGCA
151 GGTTCGAGC CCTGTCAACC AGCTCCGAGC CGGCAGCGAA ACCTGAAAGT
201 GACCCCTGTG AAAATGAAGC TGTCGCCCA GAATCACCA ACCGGAACCC
251 CCGGAACCTG GAGCTTTTGT CTGTAGCCAG GAAAGAGCGG GGCTGGCGGA
301 CGGTGTTTCC CTCCCGTGAG TTCTGGCACA GGTTGCGAGT TATAAGGACT
351 CAGCATCATG TAGAAGCACT TGTGGAGCAT CAGAATGGCA AGGTTGTGGT
401 TTCGGCTCCT ACTCGTGAGT GGGCTATTAA AAAGCACCTT TATAGTACCA
451 GAAATGTGGT GGCTTGTGAG AGTATAGGAC GAGTGCTGGC ACAGAGATGC
501 TTAGAGGCGG GAATCAACTT CATGGTCTAC CAACCAACCC CGTGGGAGGC
551 AGCCTCAGAC TCGATGAAAC GACTACAAAG TGCCATGACA GAAGGTGGTG
601 TGGTTCTACG GGAACCTCAG AGAATCTATG AATAAATGGA AGCATTAAAT
651 GTTTTGAACA TGTAATATA AATCTGTGAG CCACTACAGC CATCAAAAGA
701 GAGCATCTGG AAGAACAGCC AGCTTGAAG TTTTACAGCA ATAATGTTGC
751 AGTGGAAATAT TATTTGTAGT TAAGGTCATC CTCCTCCCCT TTCTGTTTTT
801 TTAATCAAG AACTACGTTT TGCCCTCTC TTGGGCTTCA GAAGCATCTA
851 AGAAAAGCAG TCATCAATTA TAATTAACCT TCAAAGGGA AGTCAGAAAT
901 TGTTTATAAA TTACAAATA AAGGCATATT ATGAACTCTA AAAAAAAAAA
951 AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 93 bp to 632 bp; peptide length: 180
 Category: similarity to known protein
 Classification: unset

```

1 MALRSRFWGL FSVCRNPGCR FAALSTSSEP AAKPEVDPVE NEAVAPEFTN
51 RNPRNLELLS VARKERGWRV VFPSREFWHR LRVIRTQHHV EALVEHONGK
101 VVVSASTREW AIKKHLYSTR NVVACESIGR VLAQRCLGAG INFMVYQPTP
151 WEAASDSMKR LQSAMTEGGV VLREPQRIYE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2h1, frame 3

229

DKFZphfbr2_2h10

group: brain derived

DKFZphfbr2_2h10 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2176 bp

Poly A stretch at pos. 2161, polyadenylation signal at pos. 2143

```
1 TGGGGAGTAT TCTAATTATA TTTTATATTT AATAAATTAT TTTTCTATTT
51 CTTTGTATATA TTAAGTTGCA CACTTGTTTC TTTTATCCAG AAAGTTTAGT
101 ATAATAAAAA TAGTTTAAAG ATTAACCTGTG AATGTAAAGG AAAAGTATTA
151 TTAATTATTT CAGGAAATTG CAAGACCTAA CATGGCTGAA AGAGAAACAG
201 AAACATCAAA TTCTGAAAGT AAACAAGATA AAGCTGCTTC TTCAAAAGAA
251 AAAAATGGAT GTAATGCAAA TTCATTTGAA GGCTCATCAA CAACAAAAAG
301 TGAAGAAAGC ATAACAGTTT CAGATAAGGA AAATGAAACC TGTCTTGCAG
351 ACCAGGAAAC TGGCTCAAAA AACATCGTCA GTTGTGATTC AAATATTGGT
401 GCAGATAAAG TGGAAAAGAA AAAACAAATA CAACACGTTT GTCAGGAAAT
451 GGAGTTGAAG ATGTGCCAGA GTTCAGAAAA CATAATCTTA TCTGATCAGA
501 TTAAGATCA CAACTCCAGT GAAGCCAGAT TTTCTTCAA GAATATTAAG
551 GATTTCGAT TAGCATCAGA TAATGTAAGC ATTGATCAGT TTTTGAGAAA
601 AAGACATGAA CCTGAATCTG TTAGTTCTGA TGTTAGCCAG CAAGGCAGTA
651 TTCATTTTGA ACCTCTGACT CCATCCGAGG TACTTGAGTA TGAAGCCACA
701 GAGATTCCTC AGAAAGGTAG TGGTGATCCT TCAGCCAAGA CTGATGAAGT
751 AGTGCTGAT CAAACAGATG ACATTCCTGG AGGAAATAAC CCTAGCACAA
801 CAGAGGCCAAC AGTAGACCTG GAAGATGAAA AAGAAAGAAG TTGAAATTAG
851 TCATTTTAAAG TTTCACTGTA CCAACGATAA GGGCATTGTT AACAGTGCTA
901 TCAGGTGAGC TCAGTGGTGC TGTTGTAGGT TCAGAAATGG AAATATGTAA
951 GGGAGGTCAC ACATACACTT TACCTGTATG TTCAACCTAT GTTATCAAAC
1001 AAACCAATTC ACCAATAATA GCATGATTAG TAGGGATTCC CAAAAAGTTT
1051 TTAACAAACAC GAACAGGATT TTAATGATAA TTAATTTTGC AGTGGAAAGG
1101 TCTCATTTAA TGGTTTTCAA GGAATGGGA TTTGGTTGCT GACATGAATT
1151 GATGATATTA GTAATATTTA TAAAGCCTTT CAAACTTCCA TCAATCCTAA
1201 GCTAAAAATC TTTATTACCT GTATATCCTT TTCAGTTAAC TGAGAGGAAG
1251 GGATTGGGAA ACCATGTACT TTTGGGGAGT AATTGATTAA AAACAATGGC
1301 TGATTGGCAT TGTTAATGAA GGCTTTTATTT GTGAGGATGA TGCTGGTAAA
1351 TGGAGCATGC TTAGAGTACT AAATTGATCT AATGAGAATT TGGATGAACA
1401 TAACTTAAT TTTGGATTTA ATATAACATT CCAGTCAGAC GCATGTAAC
1451 AGAATATTTG AATCTTTGTA CCTCCATACA AGTGTTAGCC TGCCAGGCTG
1501 TAAGCTTACC TTAATTAAAC TTTCAGTGAA AGTGGAATTA TTAAGATATA
1551 AATTATATTT TGTGCTTTT GTCAGTGTGT AAGCTGTGTA GAAATTTCTT
1601 GATGTATTAG TTGTATTAAT GTAAAGTAGA AACCCATTGT TGAACTCCT
1651 GTAGCTATTA TGCTTTTAAT ATTGTTTTAA TGTCTTCCT TAGAAATAGG
1701 CCCATAAAAA TGGTCTGGAA GCCAAACCAA AGTATGGTAT AATGTAGATA
1751 TTGTAAAGCA GTAAACTGAA AACATGTCCT GGCATGTATT CAGCCATGTT
1801 TAAGTGACTT TTCTGTAATT GTAAAAATAA AACTTCAAAT GGGACCTAAA
1851 ACAGTGATGT AAAAGAAGCT GTTTTGGAAG TTTAGCCTAA TTTATCTATA
1901 AGATGGCTGC TAAATTGATT TTTCAGTTCT TTTTATCATC TAAATATATA
1951 TAGATATAGA AATGAATAAT ATGAAGAACA GTAGTTTGCT TTGAAATACT
2001 AATAAACTTT TATTTAAGAT GCTTCATTTT TACTTCTTAA AACGTGCTTT
2051 GGATTCTTAA ATTTTGTTC ACTGAATGTT CAATGTTTAA AATGGCGATT
2101 AAAATACTCT GCTGTATATA GTAGTTTTTG AGTAAATATT TGCAATAAAA
2151 ATCTGCCCCC GAAAAAATAA AAAAAA
```

BLAST Results

Entry G35287 from database EMBL:
human STS SHGC-37375.

Score = 2163, P = 2.8e-91, identities = 437/441

PS00008	34->40	MYRISTYL	PDOC00008
PS00008	201->207	MYRISTYL	PDOC00008

Pfam for DKF2phfbr2_2h10.2

HMM_NAME	TNFR/NGFR cysteine-rich region		
HMM	*CpeG.tYtD.WNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*		
	+E+ T +D +N ++C E G+ + +C+++ +		
Query	40	SEESITVSDKEN--ETC--LADQET--GSKNIVSCDSNIGADK	76

DKFZphfbr2_2i17

group: intracellular transport and trafficking

DKFZphfbr2_2i17.3 encodes a novel 201 amino acid putative GTP-binding protein related to Rab1B.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory(biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes. Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells.

The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

Medline

96245776: Intracellular transport and maturation of nascent low density lipoprotein receptor is blocked by mutation in the Ras-related GTP-binding protein, RAB1B

strong similarity to rab1

complete cDNA, complete cds, start at 47, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1985 bp

Poly A stretch at pos. 1901, polyadenylation signal at pos. 1859

```

1 GGGAGCAGAG TCGACTGGGA GCGACCGAGC GGGCCGCCGC CGCCGCCATG
51 AACCCCGAAT ATGACTACCT GTTTAAGCTG CTTTGTATG GCGACTCAGG
101 CGTGGGCAAG TCATGCCCTGC TCCTGCGGTT TGCTGATGAC ACGTACACAG
151 AGAGCTACAT CAGCACCATC GGGGTGGACT TCAAGATCCG AACCATCGAG
201 CTGGATGGCA AAACATATCAA ACTTCAGATC TGGGACACAG CGGGCCAGGA
251 ACGGTTCGCG ACCATCACTT CCAGCTACTA CCGGGGGGCT CATGGCATCA
301 TCGTGGTGTA TGACGTCACT GACCAGGAAT CCTACGCCAA CGTGAAGCAG
351 TGGCTGCAGG AGATTGACCG CTATGCCAGC GAGAACGTCATAAGCTCCT
401 GGTGGGCAAC AAGAGCGACC TCACCACCAA GAAGGTGGTG GACAACACCA
451 CAGCCAAGGA GTTTGCAGAC TCTCTGGGCA TCCCTTCTT GGAGACGAGC
501 GCCAAGAAAT CCACCAATGT CGAGCAGGCG TTCAAGACCA TGGCTGCTGA
551 AATCAAAAAG CCGATGGGGC CTGGAGCAGC CTCTGGGGGC GAGCGGCCCA
601 ATCTCAAGAT CGACAGCACC CCTGTAAAGC CGGCTGGCGG TGGCTGTTGC
651 TAGGAGGGGC ACATGGAGTG GGACAGGAGG GGGCACCTTC TCCAGATGAT
701 GTCCCTGGAG GGGGGAGGAG GTACCTCCCT CTCCCTCTCC TGGGGCATT
751 GAGTCTGTGG CTTTGGGGTG TCCTGGGCTC CCCATCTCCT TCTGGCCCAT
801 CTGCCTGCTG CCCTGAGCCC CGGTCTGTGC AGGGTCCCTA AGGGAGGACA
851 CTCAGGGCCT GTGGCCAGGC AGGGCGGAGG CCTGCTGTGC AGTTGCCTCT
901 AGGTGACTTT CCAAGATGCC CCCCTACACA CCTTTCTTTG GAACGAGGGC
951 TCTTCTGTCG GTGTCCCTCC CACCCCATG TATGCTGCAC TGGGTCTCT
1001 CCTTCTTCTT CCTGCTGTCC TGCCCAAGAA CTGAGGGTCT CCCCAGGCTC
1051 TACTGCCCTG GCTGCAGTCA GTGCCAGGG CGAGGAATGT GGCCAGGGGA
1101 TCCAGGACCT GGGATCCAGG GCCCTGGGCT GGACCTCAGG ACAGGCATGG
1151 AGGCCACAGG GGCCACGAGC CCCACCTTT CCTCTCCCA CTGCCTCTC
1201 TCCCTTCTTA CACTCCAGC TCGAGCCGTC CAGCTGCGGT GGGATCTGAG
1251 TATATCTAGG GCGGGTGGGC GGGTAGCAGT GCTGGGCTG TGTCTTGAGC
1301 CTGGAGGGAG ACTGCTCCTG CCGCCCTCTG CCCTGCCGGA GACAGACCCA
1351 TGCGCTGCCT GCCCACCCTG CCCCTTTGTC CCCATGTCAG GCGGAGGCGG
1401 AAGGCCACAC GTGCCAGAGG CTGGGCACCA GCCTTAACCC TCACTCTGCT
1451 AGCACCTCCT CCCTTTCCCC AAGGTAGCAC ATCTGGCTCA CTCCCCACTC
1501 CGTCTCTGGA GCCCACCAGG GAAGGCCCTC ATCCCTGCC GCTACTTCTC
1551 TGGGGAATGT GGGTTCCATC CAGGATTGGG GGCCCTCTCT CTCACCCACT
1601 CTGCACCCAG GATCCTAGTC CCCTGCCCTC TGGCACAGCT GCTTCCTGCA
1651 AGAAAGCAAG TCTTTGGTCT CCCTGAGAAG CCATGTCCCT CGTGCTGTCT
1701 CTTGCCTGTC CCACCTGTGC CTTGCCCTCC AGCTTGATTT TAAGTCCCTG
1751 GGCTGCCCCC TTGGGGTGCC CCCCCTCCC AGGTTCCCT CTGGTGTCAT
1801 GTCAGGCATT TTGCAAGGAA AAGCCACTTG GGGAAAGATG GAAAAGGACA
1851 AAAAAAATTA ATAAATTTC ATTGGCCCTC GGGTGAGCTG AGGGTTTTTG
1901 CAAGGAAAAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
1951 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA

```


BLAST Results

No BLAST result

Medline entries

91115900:

A family of ras-like GTP-binding proteins expressed in electromotor neurons.

Peptide information for frame 3

ORF from 48 bp to 650 bp; peptide length: 201
Category: strong similarity to known protein

```

1 MNPEYDYLFK LLLIGDSGVG KSCLLRFAD DTYTESYIST IGVDKIRTI
51 ELDGKTIKIQ IWDTAGQERF RTITSSYYRG AHGIIVVYDV TDQESYANVK
101 QWLQEIADRYA SENVNKLLVG NKSDLTTKKV VDNNTAKEFA DSLGIPFLET
151 SAKNATNVEQ AFMTMAAEIK KRMGPGAASG GERPNLKIDS TPVKPAGGGC
201 C

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2i17, frame 3

SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B., N = 1, Score = 1023, P = 2.7e-103

PIR:S06147 GTP-binding protein rab1B - rat, N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:RAB1_DISOM RAS-RELATED PROTEIN ORAB-1., N = 1, Score = 967, P = 2.4e-97

PIR:TVHUYP GTP-binding protein Rab1 - human, N = 1, Score = 966, P = 3e-97

>SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B.
Length = 201

HSPs:

Score = 1023 (153.5 bits), Expect = 2.7e-103, P = 2.7e-103
Identities = 197/201 (98%), Positives = 199/201 (99%)

```

Query:      1 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ 60
             MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ
Sbjct:      1 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ 60

Query:      61 IWDTAGQERFRTITSSYYRGAGHIIVVYDVTDQESYANVKQWLQEIADRYASENVNKL
             IWDTAGQERFRTITSSYYRGAGHIIVVYDVTDQESYANVKQWLQEIADRYASENVNKL
Sbjct:      61 IWDTAGQERFRTITSSYYRGAGHIIVVYDVTDQESYANVKQWLQEIADRYASENVNKL 120

Query:      121 NKSDLTTKKVVDNNTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180
             NKSDLTTKKVVDNNTAKEFADSLG+PFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
Sbjct:      121 NKSDLTTKKVVDNNTAKEFADSLGVPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180

Query:      181 GERPNLKIDSTPVKPAAGGGCC 201
             GERPNLKIDSTPVK A GGCC
Sbjct:      181 GERPNLKIDSTPVKSASGGCC 201

```

Pedant information for DKFZphfbr2_2i17, frame 3

Report for DKFZphfbr2_2i17.3

[LENGTH] 201

[MW] 22171.25
 [PI] 5.56
 [HOMOL] SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B. 1e-112
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae, YGL210w] 1e-44
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c] 1e-30
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 1e-16
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 4e-10
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 9e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPL051w] 3e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-05
 [BLOCKS] BL01019A ADP-ribosylation factors family proteins
 [BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
 [SCOP] dlpk_ 3.25.1.3.1 cH-p21 Ras protein (human (Homo sapiens)) 2e-41
 [SCOP] dlguaa_ 3.25.1.3.10 Rap1A (Human (Homo sapiens)) 5e-60
 [SCOP] dlrrga_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattus 2e-30
 [SCOP] dlhura_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Homo 2e-33
 [PIRKW] nucleus 1e-21
 [PIRKW] membrane trafficking 1e-110
 [PIRKW] oncogene 1e-25
 [PIRKW] endoplasmic reticulum 1e-105
 [PIRKW] phosphoprotein 1e-105
 [PIRKW] glycoprotein 3e-25
 [PIRKW] prenylated cysteine 1e-110
 [PIRKW] signal transduction 4e-23
 [PIRKW] transforming protein 1e-105
 [PIRKW] purine nucleotide binding 2e-24
 [PIRKW] alternative splicing 5e-26
 [PIRKW] P-loop 1e-110
 [PIRKW] lipoprotein 1e-110
 [PIRKW] proto-oncogene 3e-27
 [PIRKW] methylated carboxyl end 3e-27
 [PIRKW] hydrolase 7e-25
 [PIRKW] membrane protein 1e-105
 [PIRKW] GTP binding 1e-110
 [PIRKW] thiolester bond 5e-76
 [PIRKW] Golgi apparatus 1e-105
 [SUPFAM] ras transforming protein 1e-110
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 5
 [PROSITE] SIGMA54_INTERACT_1 1
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 3
 [PFAM] Ras family (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D


```

SEQ      MNPEYDYLFKLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKIQ
221p-    .....EEEEEEETTTCHHHHHHHHHHCCCCCCCCCTTTEEEE-EEEEETEEEEEE

SEQ      IWDTAGQERFRTITSSYYRGAGHGIIVVYDVTQESYANVKQWLQEI DRYASENVNKL LVG
221p-    EEECTTTTTTCGGGHHHHHHHCCCEEEEEETTBHHHHHHHHHHHHHHHHHHHHHTTTTCEEEEE

SEQ      NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
221p-    ETTTTCCTCC-CCCHHHHHHHHHHCCCEEEEEETTTTTHHHHHHHHHHHHHHH.....

SEQ      GERPNLKIDSTPVKPAAGGCC
221p-    .....

```

Prosites for DKFZphfbr2_2i17.3

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	133->137	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	135->138	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00007	27->34	TYR_PHOSPHO_SITE	PDOC00007
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00017	15->23	ATP_GTP_A	PDOC00017
PS00675	11->25	SIGMA54_INTERACT_1	PDOC00579

Pfam for DKFZphfbr2_2i17.3

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVIGDSGVGKSCLLIREFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK		
Query	10	KL+LIGDSGVGKSCLL+RF++++E+YI+TIGVDF+++TIE+DGKTIK	58
HMM	LQIWDTAGQERYRsmRPMYYRGAMGFMLVYDITNRqSFENIrNWweEI+r		
Query	59	LQIWDTAGQER+R++++YYRGA+G+++VYD+T+++S+ N+++W++EI+R	108
HMM	HCDrDENVPIMLVGNKCDLEDQROVStEEGQeFAREWGAIPFMETSAKTN		
Query	109	+++ ENV ++LVGNK+DL +++V+ +++EFA+++G IPF+ETSAK++	155
HMM	INVEEAFMEIvReIlqrMqe.q.NqteNinidQpsrnrk...rCCCIM*		
Query	156	+NVE+AFM+++ EI++RM+ +++E +N++ +S++ K +CC	201

DKF2phfbr2_2k19

group: brain derived

DKF2phfbr2_2k19 encodes a novel 303 amino acid protein with similarity to human KIAA0378 product.

The protein contains a leucine zipper, which can mediate protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to KIAA0378

encoded by the genomic clones HS147M19/HS608E8

Sequenced by Qiagen

Locus: unknown

Insert length: 1931 bp

Poly A stretch at pos. 1866, no polyadenylation signal found

```
1 GGGGGGGGCG CGCGGTGACA GCGCGGGGTT GGGCGCGTGG GACCCAGGGG
51 GCGACAGAGG CAGCAGCAGC CCGAGGCGCTG AGGAGAGGAG ACCGGCGGGC
101 GCGGCAATGC TGGAGACCCT TCGCGAGCGG CTGCTGAGCG TGCAGCAGGA
151 TTTCACCTCC GGGCTGAAGA CTTTAAGTGA CAAGTCAAGA GAAGCAAAAG
201 TGAAAAGCAA ACCCAGGACT GTTCCATTTT TGCCAAAGTA CTCTGCTGGA
251 TTAGAATTAC TTAGCAGGTA TGAGGATACA TGGGCTGCAC TTCACAGAAG
301 AGCCAAAGAC TGTGCAAGTG CTGGAGAGCT GGTGGATAGC GAGGTGGTCA
351 TGCTTTCTGC GCACTGGGAG AAGAAAAAGA CAAGCCTCGT GGAGCTGCAA
401 GAGCAGCTCC AGCAGCTCCC AGCTTTAATC GCAGACTTAG AATCCATGAC
451 AGCAAATCTG ACTCATTTAG AGGCGAGTTT TGAGGAGGTA GAGAACAACC
501 TGCTGCATCT GGAAGACTTA TGTGGGCAGT GTGAATTAGA AAGATGCAAA
551 CATATGCAGT CCCAGCAACT GGAGAATTAC AAGAAAAATA AGAGGAAGGA
601 ACTTGAAACC TTCAAAGCTG AACTAGATGC AGAGCACGCC CAGAAGGTCC
651 TGGAAATGGA GCACACCCAG CAAATGAAGC TGAAGGAGCG GCAGAAGTTT
701 TTTGAGGAAG CCTTCCAGCA GGACATGGAG CAGTACCTGT CCACTGGCTA
751 CCTGCAGATT GCAGAGCGGC GAGAGCCCAT AGGCAGCATG TCATCCATGG
801 AAGTGAACGT GGACATGCTG GAGCAGATGG TCCTGATGGA CATATCGGAC
851 CAGGAGGCCC TGGACGTCTT CCTGAACTCT GGAGGAGAAG AGAACACTGT
901 GCTGTCCCCC GCCTTAGGTA GGGTTGACAA ACTTGCATTA GCTGAACCA
951 GGCAGTATCG ATGCCACTCC CCTCCAAAGG TGAGACGTGA GAACCATCTG
1001 CCAGTCACTT ACGCATAAAC CCCCAGCTC ACAGCCAGCT CCTGGCTCCC
1051 TAACCCACAG GTTCCACACG GCTGTGTGGC AGCTGCAACA GTGGTGTGGT
1101 TCCGTCAATG ATTCTTCTCA AAGATTTGAC ATGCTCCACT CCGGTAACCT
1151 TGGTGAGTTG AGAGCTTTCT TGTTTGTTT CCCTCCTTTA CCATCCAGAA
1201 ATCCATTGTA GTCTGCTCCT TGTGGTTAAG GACTGGCGTT TGCAGGGAGG
1251 TGCGGACTCT CCTGCGGGGC TCACGGGAAA CTCTTCCCTC TTCGTGCGAC
1301 AGGCATTTAG GGGCGTGCCT GCCATGGGCA AAGCCATGGT GTGTGTTTCA
1351 CTCTTGGCCT GTGTGTGAAA CTTAGTTGCA CTTCACTTCC TTTCATCCCT
1401 TCACAAAATT TTGTTTCA CAATGTCAGC AAATATGGGC TGAGGTGCCA
1451 GACCTGTACC TGGGCTTGGT GCGTTTCAA TTTCAAGACA GTTCTTTGGG
1501 CTGGGTCAA GCAAAGCTCA GTCGTCCCAG CAGCACCTCA GCCATCTGTA
1551 GAAGGTTCTA CCATTACCAC GGTTCAGCT TCCTCTAAAC TTCTCACCCG
1601 CTTCTCCTGG CAATCTGTCA GAACGGTGTG ATCCTGGGGA AGAGAAGGAG
1651 CTTGGGTGCA TTTGCCCTCA TCCTGAGAAG GCCAGAATAC TGGAGACCA
1701 CGTGAACCCCT CACCCAGAGT CAGGGGAAGA TTTAGAAACA GTGACACCTG
1751 CATATAGAA TTTGATTCTT TGAAGAGCCT ATTTAGTTCC ATAAAAATTG
1801 AGAACTGCTG AAGGTCAGTA ATTCCGACTT TCTCAGCAGT GGTGTCTCTG
1851 AATTACTGCA AAGGGTAAAA AAAAAAATA AAAAAACTTA TCGATACCGT
1901 CGACCTCGAT GATGATGATG ATGATGTCGA C
```

BLAST Results

Entry HS147M19 from database EMBL:
Homo sapiens DNA sequence from PAC 147M19 on chromosome 6p22.1-22.3.
Contains an unknown gene, ESTs and GSSs.
Score = 5540, P = 4.1e-275, identities = 1114/1120
3 exons 592-1884

Entry HS608E8 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 608E8
Score = 797, P = 1.2e-78, identities = 161/163

6 exons 1-592

Medline entries

90294724:

The involucrin gene of the gibbon: The middle region shared by the hominoids

Peptide information for frame 2

ORF from 107 bp to 1015 bp; peptide length: 303
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (97-119)

```

1 MLETLRERLL SVQDDFTSGL KTLSDKSREA KVKS KPRTVP FLPKYSAGLE
51 LLSRYEDTWA ALHRRADCA SAGELVDSEV VMLSAHWEKK KTSLEVELQEQ
101 LQQLPALIAD LESMTANLTH LEASFEEVEN NLLHLEDLCG QCCLERCKHM
151 QSQQLENYKK NKRKELETFK AELDAEHAQK VLEMEHTQOM KLKERQKFFE
201 EAFQDDMEQY LSTGYLQIAE RREPIGSMSS MEVNVDMLEQ MVLMDISDQE
251 ALDVFLNSGG EENTVLSPAL GRVDKLALAE PGQYRCHSPP KVRRENHLPV
301 TYA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2k19, frame 2

TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds., N = 1, Score = 137, P = 4.8e-06

PIR:I37037 involucrin - common gibbon, N = 1, Score = 124, P = 7.4e-05

PIR:A57013 early endosome antigen 1 - human, N = 1, Score = 128, P = 9.5e-05

>TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds.

Length = 808

HSPs:

Score = 137 (20.6 bits), Expect = 4.8e-06, P = 4.8e-06
 Identities = 59/222 (26%), Positives = 103/222 (46%)

```

Query:      2 LETLRERLLSVQDDFTSGLKTL---SDKSREAKVKS-KPRTVPFLPKYSAGLELLSRYED 57
             L TL E L S ++   LK      D+ R +++S +       K +A   L+ E
Sbjct:    434 LATLEAAL-SEKERIIERLKEQRRERDDRRERLEEIESFRKENKDLKEKVNALQAELETEKES 492

```

```

Query:      58 TWAALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPALIADLESMTAN 117
             + L A ASAG DS++ L E+KK +L+ QL++ I D M
Sbjct:    493 SLIDLKEHASSLASAGLKRDSKLKSLAIEQKKEECKLEAQLKKAHN-IEDDSRMNPE 551

```

```

Query:     118 LTHLEASFEEVENNLLHLEDLCG--QCCLERCKHMQSQQLLENYKKNKRK---ELETFAE 172
             +++++ + D CG Q E++R + .++EN K +K K ELE+
Sbjct:     552 FAD---QIKQLDKEASYRDECGKAQAEVDRLLEIL-KEVENEKNDKDKKIAELESITLR 607

```

```

Query:     173 LDAEHAQKVLEMEHTQOMKLKERQKFFEEAFQDDMEQYLLSTGYLQIAE 220
             + +KV ++H QQ++ K+ + EE +++ ++ +LQI E
Sbjct:     608 HMKDQNKKVANLKHNNQQLKKNNAQLLEEVRRREDSMADNSQHLQIEE 655

```

Score = 100 (15.0 bits), Expect = 6.2e-02, P = 6.0e-02
 Identities = 44/156 (28%), Positives = 76/156 (48%)

```

Query:      57 DTWAALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPAL-IADLESMT 115
             D A+ +R +C A VD + +L E +K + +L+ L + D
Sbjct:     560 DKEASYR--DECGKAQAEVDRLLEILK-EVENEKNDKDKKIAELESITLRHMKDQNKKV 616

```

```

Query:     116 ANLTHLEASFEEVENNLLHLEDLCGQCE--LERCKHMQSQQLLENYKKNKRKELETFKAE 173

```


ANL H + E+ +N L LE++ + + + +H+Q ++L N + R+EL+ KA L
 Sbjct: 617 ANLKHNQ-QLEKKKNAQL-LEEVRREDSDMADNSQHLQIEELMNALEKTRQELDATKARL 674
 Query: 174 DAEHAQKVLEME-HTQQMKLKERQKFFEEAFQDMEQYLS 212
 A Q + E E H +++ ER+K EE + E L+
 Sbjct: 675 -ASTQQSLAEKEAHLANLRI-ERRKQLEEILEMKQEALLA 712

Pedant information for DKFZphfbr2_2k19, frame 2

Report for DKFZphfbr2_2k19.2

[LENGTH]	303
[MW]	34814.78
[pI]	5.23
[PROSITE]	LEUCINE_ZIPPER 1
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 3.63 %
[KW]	COILED_COIL 14.52 %

SEQ	MLETLRERLLSVQQDFTSGLKTLSDKSREAKVKSKPRTVPFLPKYSAGLELLSRVEDTWA
SEG
PRD	ccchhhhhhhhhhhccccccchhhhhhhhhhhccccccchhhhhhhhhhhchhhh
COILS

SEQ	ALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLVLEQEQLOQLPALIADLESMTANLTH
SEGxxxxxxxxxxxxx.....
PRD	hhhhhhhhchhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILSCCCCCCCCCCCCCCCCCCCCCCCC

SEQ	LEASFEEVENNLLHLEDLCGQCELERCKHMQSQOLENYKKNRKELETFAELDAEHAQK
SEG
PRD	hhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS	CCCCCCCCCCCCCCCCCCCC.....

SEQ	VLEMEHTQQMKLKERQKFFEEAFQDMEQYLSGTGYLQIAERREPIGSMSSMEVNVDMLEQ
SEG
PRD	hh
COILS

SEQ	MVLMDISDQEALDVFLNSGGEENTVLSPALGRVDKLALAEQYRCHSPPKVRRENHLPV
SEG
PRD	hhhhhhchhhhhhhhhccccceeeccccccccceeeccccccccceeecccccc
COILS

SEQ	TYA
SEG	...
PRD	ccc
COILS	...

Prosites for DKFZphfbr2_2k19.2

PS00029 97->119 LEUCINE_ZIPPER PDOC00029

(No Pfam data available for DKFZphfbr2_2k19.2)

DKFZphfbr2_2k14

group: cell cycle

DKFZphfbr2_2k14 encodes a novel 335 amino acid protein with strong similarity to *rattus rattus* IAG2 "implantation-associated protein" and the human N33 tumour-suppressor gene.

Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

The new protein can find application in modulating/blocking the cell cycle and in the therapy of tumours.

strong similarity to human N33 tumor suppressor gene

complete cDNA, complete cds, EST hits,
potential start at Bp 30 matches kozak consensus ANCatgG
potential transmembran protein (4 TM)
similarity to yeast OST3p (oligosaccharyltransferase gamma chain)

Sequenced by Qiagen

Locus: unknown

Insert length: 2241 bp
Poly A stretch at pos. 2221, no polyadenylation signal found

```

1 TGGGACTTAT AGAAGGGAGA GGAGCGAACA TGGCAGCGCG TTGGCGGTTT
51 TGGTGTGTCT CTGTGACCAT GGTGGTGGCG CTGCTCATCG TTTGCGACGT
101 TCCCTCAGCC TCTGCCAAA GAAAGAAGGA GATGGTGTTA TCAGAAAAGG
151 TTAGTCAGCT GATGGAATGG ACTAACAAAA GACCTGTAAT AAGAATGAAT
201 GGAGACAAGT TCCGTCGCCT TGTGAAAGCC CCACCGAGAA ATTACTCCGT
251 TATCGTCATG TTCCTGCTC TCCAACTGCA TAGACAGTGT GTCGTTTGCA
301 AGCAAGCTGA TGAAGAATTC CAGATCCTGG CAAACTCCTG GCGATACTCC
351 AGTGCAATCA CCAACAGGAT ATTTTGTGCC ATGGTGGATT TTGATGAAGG
401 CTCTGATGTA TTTCAGATGC TAAACATGAA TTCAGCTCCA ACTTTCATCA
451 ACTTTCCTGC AAAAGGGAAA CCCAACGGG GTGATACATA TGAGTTACAG
501 GTGCGGGGTT TTTCAGCTGA GCAGATTGCC CGGTGGATCG CCGACAGAAC
551 TGATGTCAT ATTAGAGTGA TTAGACCCCC AAATTATGCT GGTCCCCTTA
601 TGTGGGATT GCTTTTGGCT GTTATTGGTG GACTTGTGTA TCTTCGAAGA
651 AGTAATATGG AATTTCCTTT TAATAAACT GGATGGGCTT TTGCAGCTTT
701 GTGTTTGTG CTGTCTATGA CATCTGGTCA AATGTGGAAC CATATAAGAG
751 GACCACCAT TGCCCATAG AATCCCAACA CGGGACATGT GAATTATATC
801 CATGGAAGCA TCAAGCCCA GTTTGTAGCT GAAACACACA TTGTCTTCT
851 GTTTAATGGT GGAGTTACCT TAGGAATGGT GCTTTTGTGT GAAGCTGCTA
901 CCTCTGACAT GGATATTGGA AAGCGAAAAG TAATGTGTGT GGCTGGTATT
951 GGACTTGTG TATTATTCTT CAGTTGGATG CTCTCTATT TTAGATCTAA
1001 ATATCATGGC TACCCATACA GCTTCTGAT GAGTTAAAAA GGTCCCAGAG
1051 ATATATAGAC ACTGGAGTAC TGGAAATTGA AAAACGAAAA TCGTGTGTGT
1101 TTGAAAAGAA GAATGCAACT TGTATATTCT GTATTACCTC TTTTTCCTAA
1151 GTGATTAAA TAGTTAATCA TTTAACCAAA GAAGATGTGT AGTGCCTTAA
1201 CAAGCAATCC TCTGTCAAAA TCTGAGGTAT TTGAAAATAA TTATCCTCTT
1251 AACCTTCTCT TCCCAGTGAA CTTTATGGAA CATTTAATTT AGTACAATTA
1301 AGTATATTAT AAAAATTGTA AACTACTAC TTTGTTTGTAG TTAGAACAAA
1351 GCTCAAACT ACTTTAGTTA ACTTGGTCAT CTGATCTTAT ATTGCCTTAT
1401 CCAAAGATGG GGAAAGTAAG TCCTGACCAG GTGTTCCAC ATATGCCTGT
1451 TACAGATAAC TACATTAGGA ATTCATTCTT AGCTTCTTCA TCTTGTGTG
1501 GATGTGTATA CTTACGCAT CTTTCTTTT GAGTAGAGAA ATTATGTGTG
1551 TCATGTGGTC TTCTGAAAAT GGAACACCAT TCTTCAGAGC ACACGTCTAG
1601 CCTCAGCAA GACAGTTGTT TCTCCTCCTC CTTGCATATT TCCTACTGCG
1651 CTCCAGCCTG AGTGATAGAG TGAGACTCTG TCTCAAAAAA AAAGTATCTC
1701 TAAATACAGG ATTATAATTT CTGCTTGAGT ATGGTGTAA CTACCTTGTA
1751 TTTAGAAAAG TTTCAGATTC ATTCCATCTC CTTAGTTTTC TTTTAAGGTG
1801 ACCCATCTGT GATAAAAATA TAGCTTAGTG CTAAAAACAG TGTAACCTAT
1851 ACATGGCCTA AATGTTTCT ACAAATTAGA GTTGTCACT TATTCCATTT
1901 GTACCTAAGA GAAAAATAGG CTCAGTTAGA AAAGGACTCC CTGGCCAGGC
1951 GCAGTGACTT ACGCCTGTAA TCTCAGCACT TTGGGAGGCC AAGGCAGGCA
2001 GATCAGGAG TCAGGAGTTC GAGACCATCC TGGCCAACAT GGTGAAACCC
2051 CGTCTCTACT AAAAATATAA AAATTAGCTG GGTGTGGTGG CAGGAGCCTG
2101 TAATCCAGC TGCACAGGAG GCTGAGGCAC GAGAATCACT TGAACCTCAGG
2151 AGATGGAGGT TTCAGTGAGC CGAGATCACG CCACTGCACT CCAGCCTGGC
2201 AACAGAGCGA GACTCCATCT CAAAAA AAAA A

```


BLAST Results

No BLAST result

Medline entries

96299740:
Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22.

97243398:
Tumour-suppressor genes in prostatic oncogenesis: a positional approach.

98334474:
Concordant methylation of the ER and N33 genes in glioblastoma multiforme.

Peptide information for frame 3

ORF from 30 bp to 1034 bp; peptide length: 335
Category: strong similarity to known protein

```

1 MAARWRFVCV SVTMVVALLI VCDVPSASAQ RKKEMVLSEK VSQLMWNTK
51 RPVIRMNGDK FRRLVKAPPR NYSVIVMFTA LQLHRQCVVC KQADEEFQIL
101 ANSWRYSSAF TNRIFFAMVD FDEGSDVFQM LNMNSAPTFI NFPKKGKPKR
151 GDTYELQVRG FSAEQIARWI ADRTDVNIRV IRPPNYAGPL MLGLLAVIG
201 GLVYLRRSNM EFLFNKTGWA FAALCFVLAM TSGQMWNHIR GPPYAHKNPH
251 TGHVNYIHGS SQAQFVAETH IVLLFNGGVT LGMVLLCEAA TSDMDIGKRR
301 IMCVAGIGLV VLFFSWMLSI FRSKYHGYPY SFLMS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2k14, frame 3

TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds., N = 1, Score = 1560, P = 3.4e-160

PIR:G02297 gene N33 protein - human, N = 1, Score = 1256, P = 5.6e-128

TREMBL:HSN33S11_1 gene: "N33"; product: "N33 protein form 2"; Human N33 protein form 2 (N33) gene, exon 11 and complete cds., N = 1, Score = 1252, P = 1.5e-127

>TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds. Length = 308

HSPs:

Score = 1560 (234.1 bits), Expect = 3.4e-160, P = 3.4e-160
Identities = 295/307 (96%), Positives = 299/307 (97%)

```

Query:   29 AQRKKEMVLSEKVSQLMWNTNRPVIRMNGDKFRRLVKAPPRNYSVIVMFTALQLHRQCV 88
          AQRKKE VL EKV QLMWNTN+RPVIRMNGDKFR LVKAPPRNYSVIVMFTALQLHRQCV
Sbjct:   2 AQRKKERVLVEKVIQLMEWTNQRPVIRMNGDKFRPLVKAPPRNYSVIVMFTALQLHRQCV 61

Query:   89 VCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPKKGK 148
          VCKQADEEFQILAN WRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFP KGKP
Sbjct:   62 VCKQADEEFQILANFWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPKKGK 121

Query:   149 KRGDYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 208
          KR DTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS
Sbjct:   122 KRADTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 181

Query:   209 NMEFLFNKTGWAFALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE 268
          NMEFLFNKTGWAFALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE

```


Pedant information for DKFZphfbr2_2k14, frame 3

Report for DKFZphfbr2_2k14.3

```

[PIRKW]      glycosyltransferase 1e-12
[PIRKW]      transmembrane protein 6e-69
[PIRKW]      hexosyltransferase 1e-12
[PROSITE]    RGD      1
[PROSITE]    MYRISTYL      4
[PROSITE]    AMIDATION      1
[PROSITE]    CK2_PHOSPHO_SITE      2
[PROSITE]    PKC_PHOSPHO_SITE      4
[PROSITE]    ASN_GLYCOSYLATION      2
[KW]         SIGNAL PEPTIDE 30
[KW]         TRANSMEMBRANE 4
[KW]         LOW COMPLEXITY      5.97 %

```

```
SEQ      IMCVAGIGLVVLFFSWMLSIFRSKYHGYPYSFLMS
SEG      .....
PRD      eeeeeccccceeeehhhhhhhhhhccccc
MEM      mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm.....
```

Prosite for DKFZphfbr2 2k14.3

PS000001	71->75	ASN_GLYCOSYLATION	PDOC00001
PS000001	215->219	ASN_GLYCOSYLATION	PDOC00001
PS000005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS000005	48->51	PKC_PHOSPHO_SITE	PDOC00005

PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00008	193->199	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	278->284	MYRISTYL	PDOC00008
PS00009	296->300	AMIDATION	PDOC00009
PS00016	150->153	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_2k14.3)

DKFZphfbr2_3c18

group: nucleic acid management

DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicase and RNA-dependent ATPase
from the DEAD box family

group helicases

Summary DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with similarity to DEAD-box subfamily ATP-dependent RNA helicases. Deletion of the yeast homologue DBP5 is lethal.

strong similarity to RNA helicase and RNA-dependent ATPase from the
DEAD box family

complete cDNA, EST hits
complete cds ATG at Bp 109

Sequenced by AGOWA

Locus: /map="87.50 cR from top of Chr16 linkage group"

Insert length: 1713 bp

Poly A stretch at pos. 1696, no polyadenylation signal found

```
1 TGGGGTAGTG GGGCTGGAGC AGAGCCTGCC GCGAACCCCC GGAGCCCACG
51 ATCCCTCGTG CCATCCCTCG AATCCACCAG CACGAGCGTC CCACCCGCGC
101 CTGGGACCAT GGCCACTGAC TCATGGGCCC TGGCGGTGGA CGAGCAGGAA
151 GCTGCGGCTG AGTCGTTGAG CAACTTGCAT CTTAAGGAAG AGAAATCAA
201 ACCAGATACC AATGGTGCTG TTGTCAAGAC CAATGCCAAT GCAGAGAAGA
251 CAGATGAAGA AGAGAAAGAG GACAGAGCTG CCCAGTCCTT ACTCAACAAG
301 CTGATCAGAA GCAACCTTGT TGATAACACA AACCAAGTGG AAGTCTGCA
351 GCGGGATCCA AACTCCCTCT TGTACTCGGT GAAGTCTTTT GAAGAGCTTC
401 GGCTCCACCA GAACTTAATT GCCCAATCTC AGTCTGGTAC TGGTAAACAA
451 GCTGCCCTCG TGCTGGCCAT GCTTAGCCAA GTAGAACCTG CAAACAAATA
501 CCCCCAGTGT CTATGTCTCT CCCCACCGTA TGAGCTCGCC CTCCAAACAG
551 GAAAAAGTGT TGAACAAATG GGCAAAATTT ACCCTGAACT GAAGCTAGCT
601 TATGCTGTTC GAGGCAATAA ATTGGAAGA GGCAGAGA TCAGTGAGCA
651 GATTGTCTAT GGCACCCCTG GGAAGTGTCT GGAAGTGTCT TCCAAGCTCA
701 AGTTTCATTA TCCCAAGAAA ATCAAGGTGT TTGTTCTGGA TGAGGCTGAT
751 GTCATGATAG CCACTCAGGG CCACCAAGAT CAGAGCATCC GCATCCAGAG
801 GATGCTGCCC AGGAACTGCC AGATGCTGCT TTTCTCCGCC ACCTTTGAAG
851 ACTCTGTGTG GAACTTTGCC CAGAAAGTGG TCCCAGACCC AAACGTTATC
901 AAATGAAGC GTGAGGAAGA GACCCTGGAC ACCATCAAGC AGTACTATGT
951 CCTGTGCAGC AGCAGAGACG AGAAGTTCCA GGCCTTGTGT AACCTCTACG
1001 GGGCCATCAC CATTGCTCAA GCCATGATCT TCTGCCATAC TCGCAAAACA
1051 GCTAGTTGGC TGGCAGCAGA GCTCTCAAAA GAAGGCCACC AGGTGGCTCT
1101 GCTGAGTGGG GAGATGATGG TGGAACAGAG GGCTGCAGTG ATTGAGCGCT
1151 TCCGAGAGGG CAAAGAGAAG GTTTTGGTGA CCACCAACGT GTGTGCCCCG
1201 GGCATTGATG TTGAACAAGT GTCTGTGCTG ATCAACTTTG ATCTTCCCGT
1251 GGACAAGGAC GGGAAATCCTG ACAATGAGAC CTACCTGCAC CGGATCGGGC
1301 GCACGGGCCC CTTTGGCAAG AGGGGCCTGG CAGTGAACAT GGTGGACAGC
1351 AAGCACAGCA TGAACATCCT GAACAGAATC CAGGAGCATT TTAATAAGAA
1401 GATAGAAAGA TTGGACACAG ATGATTGGGA CGAGATTGAG AAAATAGCCA
1451 ACTGAGAAGC TCCACCAGCC ACTGATGCCA GCCCTGGCAC TGCCCTTGCA
1501 CAGGAGACAA GTGCGTTTCA GGCACAGGCC CCGACATCAC CCCAAGGACA
1551 ACGGCACAAG TAGAGAGAAA CTACCTACCT CACTTCAAAT TATGTTTGGG
1601 CTTGACAAAA ATGTATGCAA ATGATGGGGG ATGGTAGAAA AAAATTATT
1651 ACACAACCTT GGAAGATTAG GCATGAATAC ACAGAGATT ACCTTTAAAA
1701 AAAAAAAAAA AAA
```

BLAST Results

Entry G36496 from database EMBL:
 SHGC-53094 Human Homo sapiens STS cDNA.
 Length = 459
 Minus Strand HSPs:
 Score = 1693 (254.0 bits), Expect = 2.8e-70, P = 2.8e-70
 Identities = 369/387 (95%), Positives = 369/387 (95%)

Entry G44014 from database EMBLNEW:
 WIAF-3643-STH Human THudson SANGER Homo sapiens STS genomic, sequence
 tagged site.
 Score = 901, P = 2.3e-35, identities = 183/185

Medline entries

94192995:
 Gene 1994 Mar 25;140(2):171-177
 Mouse erythroid cells express multiple putative RNA helicase genes
 exhibiting
 high sequence conservation from yeast to mammals.

Peptide information for frame 1

ORF from 109 bp to 1452 bp; peptide length: 448
 Category: strong similarity to known protein

```

1 MATDSWALAV DEQEAAAESL SNLHLKEEKI KPDNGAVVK TNANAECTDE
51 EEKEDRAAQS LLNKLIRSNL VDNTNQVEVL QRDPSNPLYV VKSFEELRLP
101 QNLIAQSQSG TGKTAAFVLA MLSQVEPANK YPQCLCLSPT YELALQTGKV
151 IEQMGKFYFE LKLAYAVRGN KLERGQKISE QIVIGTPGTV LDWCSKLKFI
201 DPKKIKVFVL DEADVMIAATQ GHQDQSIRIQ RMLPRNCQML LFSATFEDSV
251 WKFAQKVVPD PNVIKLKREE ETLDTIKQYY VLCSSRDEKF QALCNLYGAI
301 TIAQAMIFCH TRKTASWLA ELSEKGHQVA LLSGEMMVEQ RAAVIERFRE
351 GKEKVLVTN VCARGIDVEQ VSVVINFDLP VDKDGNPDNE TYLHRIGRTG
401 RFGKRGGLAVN MVDSKHSMMNI LNRIQEHFNK KIERLDTDDL DEIEKIAN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_3c18, frame 1

PIR:I49731 RNA helicase - mouse, N = 2, Score = 1758, P = 3.8e-223

TREMBL:AF005239.1 gene: "Dbp80"; product: "DEAD-box helicase";
 Drosophila melanogaster DEAD-box helicase (Dbp80) mRNA, complete cds.,
 N = 2, Score = 1142, P = 1.8e-125

SWISSPROT:YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06., N =
 2, Score = 911, P = 5.5e-103

PIR:S66920 probable RNA helicase CA5/6 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 887, P = 1.9e-98

>PIR:I49731 RNA helicase - mouse
 Length = 478

HSPs:

Score = 1758 (263.8 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 338/349 (96%), Positives = 349/349 (100%)

```

Query: 100 PQNLIAQSQSGTGKTAAFVLAMLSQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYP 159
      PQNLIAQSQSGTGKTAAFVLAMLS+VEPA++YPQCLCLSPTYELALQTGKVIEQMGKF+P
Sbjct: 130 PQNLIAQSQSGTGKTAAFVLAMLSRVEPADRYPPQCLCLSPTYELALQTGKVIEQMGKFHP 189

Query: 160 ELKLAYAVRGNKLERGQKISEQIVIGTPGTVLDWCSKLKFIIDPKKIKVFVLDEADVMIAAT 219
      ELKLAYAVRGNKLERGQK+SEQIVIGTPGTVLDWCSKLKFIIDPKKIKVFVLDEADVMIAAT
Sbjct: 190 ELKLAYAVRGNKLERGQKVEQIVIGTPGTVLDWCSKLKFIIDPKKIKVFVLDEADVMIAAT 249

Query: 220 QGHQDQSIRIQRMLPRNCQMLLFSATFEDSVWKFAQKVVPDPNVIKLKREEETLDTIKQY 279

```


Sbjct: 250 QGHQDQSIRIQR++PRNCQMLLFSAFEDSVWKFQKVVPDPN+IKLKREEETLDTIKQY 309
 Query: 280 YVLCSSRDEKFOALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 339
 YVLC++R+EKFOALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE
 Sbjct: 310 YVLCNNREEKFQALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 369
 Query: 340 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKOGNPDNETYLHRIGRT 399
 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKOGNPDNETYLHRIGRT
 Sbjct: 370 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKOGNPDNETYLHRIGRT 429
 Query: 400 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 448
 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN
 Sbjct: 430 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 478

Score = 419 (62.9 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 94/136 (69%), Positives = 104/136 (76%)

Query: 1 MATDSWALAVDEQEAAAESLSNLHLKEEKIKPDTNGAVVKTNANA EKTDEEEKEDRAAQS 60
 MATDSWALAVDEQEAAA +S+S+L +KEEK K DTNG V+KT+ AEKT+EEEEKEDRAAQS
 Sbjct: 1 MATDSWALAVDEQEAAVKSMSLQIKEEKA KSDTNG-VIKTSTTAEKTEEEKEDRAAQS 59
 Query: 61 LLNKLIRSNLVNTNQVEVLQRPNSPLYSVKSFEELRL-PQNL---IAQSQSGTGKTA 116
 LLNKLIRSNLVNTNQVEVLQRP+SPLYSVKSFEELRL PQ L A + K
 Sbjct: 60 LLNKLIRSNLVNTNQVEVLQRPSSPLYSVKSFEELRLKPQLLQGVYAMGFNRPSKIQE 119
 Query: 117 FVLAMLSQVEPANKYPQ 133
 L M+ P N Q
 Sbjct: 120 NALPMLAEPQNLIAQ 136

Pedant information for DKFZphfbr2_3c18, frame 1

Report for DKFZphfbr2_3c18.1

[LENGTH] 448
 [MW] 50490.07
 [pI] 5.83
 [HOMOL] PIR:149731 RNA helicase - mouse 0.0
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 1e-102
 [FUNCAT] 04.01.04 rRNA processing [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 2e-49
 [FUNCAT] j mRNA translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 9e-48
 [FUNCAT] 04.05.03 mRNA processing (splicing) [S. cerevisiae, YDL084w] 1e-43
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 3e-39
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-35
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 9e-27
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 8e-26
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 1e-23
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 9e-08
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 03.19 recombination and DNA repair [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIR002c] 7e-04
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-64
 [PIRKW] RNA binding 1e-64
 [PIRKW] DEAD box 4e-64
 [PIRKW] transmembrane protein 3e-22
 [PIRKW] DNA binding 2e-32
 [PIRKW] ATP 1e-101
 [PIRKW] purine nucleotide binding 4e-64
 [PIRKW] P-loop 1e-101
 [PIRKW] hydrolase 4e-43
 [PIRKW] protein biosynthesis 1e-64
 [PIRKW] ATP binding 2e-35
 [SUPFAM] WW repeat homology 3e-29
 [SUPFAM] translation initiation factor eIF-4A 1e-64
 [SUPFAM] DEAD/H box helicase homology 1e-101
 [SUPFAM] DNA helicase recG 2e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-101
 [SUPFAM] ATP-dependent RNA helicase DBP1 9e-33


```

Query      159 PELKLAYAVR----GNKLERGQKISEQIVIGTPGTVLDWCSKLFIDPKK    204
HMM                IeMLVMDEADRM LD.MGFIDQIRrIMrqIPMpwnRQTMMFSATMPdeIqE
                I+++V+DEAD M+  +G +DQ RI R++P  +N Q ++FSAT+ D++ +
Query      205 IKVFVLDEADVMIATQGHQDQSIRIQRMLP--RNCQMLLFSATFEDSVWK    252
HMM                LARrFMRNPiRInIdMdElTtnEnIkQwYiyVerEMWKfdclcrLlie*
                +A ++  +P I ++++E T++ +IKQ+Y+ + + ++KF +LC+L++
Query      253 FAQKVVPDPNVIKLKREEETLD-TIKQYYVLCSSRDEKFKQALCNLYG    298

```

HMM_NAME Helicases conserved C-terminal domain

```

HMM                *EileeWLknlGirvmYIHGdMpQeERdeIMddFnnGEynVLicTDVggr
                +L+ +L+++G +V+ + G M+ E+R  ++++F++G+ +VL++T+V +R
Query      316 SWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTNNVCAR    364
HMM                GIDIPdVNVHVINYDM....PWNPEq..YIQRIGRTgRIG*
                GID+++V++VIN+D+      + NP++ Y++RIGRTGR+G
Query      365 GIDVEQVSVINFDLPVDKDGNDNETYLRIGRTGRFG    403

```

Medline

PMID: 10322435

"Unwinding RNA in : DEAD-box proteins and related families." de la Cruz J, Kressler D, Linder P

DKFZphfbr2_3f16

group: brain derived

DKFZphfbr2_3f16 encodes a novel 127 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1514 bp

Poly A stretch at pos. 1454, polyadenylation signal at pos. 1434

```

1 GGGGGGACTG GAGAAGGGAG GCGGCGGGCG AAGCGCACGT CGAGCGGGGG
51 AGCGGCGCTG CCTGTGGAGA TCCGCGGAGG CCGACAGGAT TCGTTGGCTG
101 CCGTCCCCCG TGCTGTGCAT TGGGTTAAAA ACGACAACCA ACATCAGCCA
151 TGAAGATCC AAGTCGCAGC AGTACTAGCC CAAGCATCAT CAATGAAGAT
201 GTGATTATTA ACGGTCATTC TCATGAAGAT GACAATCCAT TTGCAGAGTA
251 CATGTGGATG GAAAATGAAG AAGAATTCAA CAGACAAATA GAAGAGGAGT
301 TATGGGAAGA AGAATTTATT GAACGCTGTT TCCAAGAAAT GCTGGAAGAG
351 GAAGAAGAGC ATGAATGGTT TATTCAGCT CGAGATCTCC CACAACTAT
401 GGACCAAATC CAAGACCAGT TTAATGACCT TGTATCAGT GAAGGCTCTT
451 CTCTGGAAGA TCTTGTGGTC AAGAGCAATC TGAATCCAAA TGCAAAGGAG
501 TTTGTTCCCTG GGGTGAAGTA CGGAAATATT TGAGTAGACG GGGCCCTCTT
551 TTGGTGGATG TAGCACAAAT TCCACACTGT GAAGGCAGTA TTAGAAGACT
601 TAATTGTAAG AGCACTCTTG TCACTGTGTT ACACCTATGC ATTGCCAAAG
651 TTTTGTGTAG TCTTGCATGC TTAATAAAAG TGCTGAGACT GTTACTAAGT
701 AAAAAGCTGT CAAACATTTA CTGAAATAG AATTGGCCCC ATGGCTTGAT
751 GTGAAGACAG CAAGGAAAGA AGCACCAGTC AAGTTGTGAA CAAGCACCAA
801 ATTAAAGAC CTAACCTTA CCAATTGTC TTTTGTGAG GCTAATCTAT
851 CACTTGTTAA TGTCTAACT TTAATCAG TACATTTAAT TTGAGTTCCA
901 ACTGTTAAGC ATATTTCTCA GACTTAAAT TGATTATGTC CCCATCAAAA
951 AGAATCTCCA TTTTCTGAAG GTCTGTTAGT TAATTTGAGA TAATTTGTTA
1001 AAGGCAAGTA TGTCAATTA CTGAGGCTAC AAGTTAGTCA GCAGATGAGT
1051 GCCAGTCCAG CCTTTCCGG TATGTTATTG TTAGAAATAT TGAGTTCTAA
1101 TGTACATCT GAGGAAGTAT GTAATTTGAG AATTGTAAT TCTAAGGGAT
1151 TCACTGCATC ATAGCTATGC CTGTATGGAG TCTAACATAT GACCAATACC
1201 AACCCATAAT CCAGCTGAAC AAAGATACTG TAACATTATG ATTTGAGTGG
1251 TGCTTTTCCT TGCTTTGTTA ACCATCACGA GAGTCTGCAG CACAACCTTT
1301 AACAAAGCTA GAACAGTTT GGCTTCTTAA ACTTCATATT TGGGTAGGTT
1351 AAGCTGCCAT ACGTGTTCAG TGTGAATAGT GTTTAAGTTG AAAATATTGT
1401 AAAAAAATTA TATTTTTC AATATTTT AAAAAATTA TAATAGTAGA
1451 ACTGAAAAAA AAAAAA AAAAAA AAAAAA AAAAAA
1501 AAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 150 bp to 530 bp; peptide length: 127
Category: putative protein

1 MKDPSRSSTS PSIINEDVII NGHSHEDONP FAEYMMENE EEFNRQIEEE

51 LWEEEFIERC FQEMLEEEEE HEWFIPARDL PQTMDQIQDQ FNDLVISEGS
101 SLEDLVVKS NLPNAKEFVP GVKYGN I

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3f16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_3f16, frame 3

Report for DKFZphfbr2_3f16.3

[LENGTH] 127
[MW] 14998.41
[PI] 4.04
[BLOCKS] BL01269D
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 2
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 27.56 %

SEQ MKDPSRSSTSPSIINEDVIINGHSHEDDNPFÄEYMWENEEEFNRQIEEELWEEEFIERC
SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD ccc

SEQ FQEMLEEEEEHEWFIPARDLPQTMDQIQDQFNDLVISEGSSLEDLVVKS NLPNAKEFVP
SEGXXXXXXXXXXXX
PRD hhhhhhhhhhhhhcc

SEQ GVKYGN I
SEG
PRD ccccccc

Prosites for DKFZphfbr2_3f16.3

PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	100->104	CK2_PHOSPHO_SITE	PDOC00006
PS00008	121->127	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_3f16.3)

DKFZphfbr2_3g8

group: metabolism

DKFZphfbr2_3g8.1 encodes a novel 178 amino acid protein with similarity to yeast ARD1 protein.

In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. The new protein could be part of this or an other NAT complex.

The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

strong similarity to N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 homolog

complete cDNA, complete cds? start at Bp 40, EST hits

Sequenced by AGOWA

Locus: /map="20"

Insert length: 1030 bp

Poly A stretch at pos. 1013, no polyadenylation signal found

```

1 TGGGCTTGGC GAACGGTCTT CGGAAGCGGC GCGGCGCGCA TGACCACGCT
51 ACGGGGCTTT ACCTGCGACG ACCTGTTCCG CTTCAACAAC ATTAACCTGG
101 ATCCACTTAC AGAACTTAT GGGATTCTT TCTACCTACA ATACCTCGCC
151 CACTGGCCAG AGTATTTTAT TGTTCAGTG GCACCTGGTG GAGAATTAAT
201 GGGTTATATT ATGGGTAAAG CAGAAGGCTC AGTAGCTAGG GAAGAATGGC
251 ACGGGCACGT CACAGCTCTG TCTGTTGCCC CAGAATTCG ACGCCTTGGT
301 TTGGCTGCTA AACTTATGGA GTTACTAGAG GAGATTTTCA AAAGAAAGGG
351 TGGGTTTTTT GTGGATCTCT TTGTAAGAGT ATCTAACCAA GTTGCAGTTA
401 ACATGTACAA GCAGTTGGGC TACAGTGTAT ATAGGACGGT CATAGAGTAC
451 TATTCGGCCA GCAACGGGGA GCCTGATGAG GACGCTTATG ATATGAGGAA
501 AGCACTTTCC AGGGATACTG AGAAGAAATC CATCATACCA TTACCTCATC
551 CTGTGAGGCC TGAAGACATT GAATAACCC TGGCAGTGGT TCTTAGGCAG
601 ATACTCTAGA TGCTTTATGG ACAATATTAT TTTCATTGGA TGATTCTGGA
651 GCTCTATTAG GAGAAAAGTA ATCATTTTAG GTCTTAAAGA CTTCAAGAAA
701 ATACAGGTTA TCAATTTATT TTAAATCTCA TTGTTTCCAG TTAGCAATAT
751 CATACCTATT AAAGCTGTTT ATTGTAACAA AATTCAATCA AAAAGGCAGC
801 TAGGTCAGAA GGAACATAC CACTCTCATG GTTCATAGTA TTCACTGTAT
851 GTATGCTAGG GAAAAGACTT GCTCCAGTCT CCTCCTCAGT TCTGTGCTG
901 AGAACCACGT CTGCATATAT TTGTTTTTAA ATTTGTGATT GAACGTGTAA
951 TTGAAGCTTT AAAAGCATAT ATGAAATGTA TAAATCTAAG ATGTATAATA
1001 CATTATTGAC TCCAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HSG0101 from database EMBL:
human STS SHGC-35956.
Length = 401
Minus Strand HSPs:
Score = 1417 (212.6 bits), Expect = 9.3e-58, P = 9.3e-58
Identities = 301/311 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 40 bp to 573 bp; peptide length: 178
Category: strong similarity to known protein

```

1 MTTLRFTCD DLFRENNINL DPLTETYGIP FYLQYLAHWP EYFIVAVAPG
51 GELMGYIMGK AEGSVAREEW HGHVTALSVA PEFRRGLGLAA KLMELLEIS

```


101 ERKGGFFVDL FVRVSNQVAV NMYKQLGYSV YRTVIEYYSA SNGEPDEDAY
151 DMRKALSRDT EKKSIIPLPH PVRPEDIE

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3g8, frame 1

TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4., N = 1, Score = 475, P = 3.2e-45

SWISSPROT:ARDH_LEIDO N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG., N = 1, Score = 451, P = 1.1e-42

PIR:S69021 hypothetical protein YPR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 382, P = 2.3e-35

>TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. Length = 180

HSPs:

Score = 475 (71.3 bits), Expect = 3.2e-45, P = 3.2e-45
Identities = 96/165 (58%), Positives = 118/165 (71%)

Query: 1 MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGE--LMGYIM 58
MT R F DLF FNNINLDPLTET+ I FYL YL WP +V + + LMGYIM
Sbjct: 1 MTDTRKFKATDLFSFNNINLDPLTETFNISFYLSYLNKWPSCVQESDLSDPTLMGYIM 60

Query: 59 GKAEQSVAREEWHGHVHTALSVAPEFRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQV 118
GK+EG+ +EWH HVTAA++VAP RRLGLA +M+ LE + + FFVDLFVR SN +
Sbjct: 61 GKSEGT--GKEWHTHVTAITVAPNSRRLGLARTMDYLETVGNSENAFFVDLFVRASNAL 118

Query: 119 AVNMYKQLGYSVYRTVIEYYASNGEPDEDAYDMRKALSRDEKKSII 165
A++ YK LGYSVYR VI YYS +G+ DED++DMRK LSRD ++SI
Sbjct: 119 AIDFYKGLGYSVYRRVIGYYSNPHCK-DEDSFDMRKPLSRDVRNRESI 164

Pedant information for DKFZphfbr2_3g8, frame 1

Report for DKFZphfbr2_3g8.1

[LENGTH] 178
[MW] 20338.24
[pI] 5.06
[HOMOL] TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. 7e-47
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPR131c] 6e-37
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] r general function prediction [M. jannaschii, MJ1530] 6e-09
[PIRKW] acyltransferase 1e-12
[SUPFAM] arrest-defective protein 1 1e-12
[SUPFAM] Escherichia coli peptide N-acetyltransferase rimI 1e-07
[PROSITE] CK2_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 3
[KW] Alpha_Beta

SEQ MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGELMGYIMGK
PRD cccccccccchhhhhccccccccccccchhhhhccccceeeeeccccceeehhhh

SEQ AEGSVAREEWHGHVHTALSVAPEFRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQVAV
PRD hccccccccccccceeehhhhhhhhccchhhhhhhhhhhhhccccceeeeeccchhhhh

SEQ NMYKQLGYSVYRTVIEYYASNGEPDEDAYDMRKALSRDEKKSIIPLPHVPVRPEDIE
PRD hhhhhccccchhhhhccccccccccccchhhhhhhhhhhhhhhhhcccccccccccc

Prosite for DKFZphfbr2_3g8.1

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	100->103	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	8->12	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	141->145	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphfbr2_3g8.1)

DKF2phfbr2_312

group: brain derived

DKF2phfbr2_312 encodes a novel 589 amino acid protein with weak similarity to *S. cerevisiae* ubiquitin-like protein DSK2.

Pfam predicts for this protein similarity to the ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ubiquitin-like protein DSK2 yeast

complete cDNA, complete cds, EST hits
Dsk2p is involved in spindle pole body SPB duplication, SPB = centomer
strong similarity to HRIHFB2157 human mRNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2978 bp
Poly A stretch at pos. 2958, polyadenylation signal at pos. 2924

```
1 GGGGGGAGGA AGCGGTGGCT GCTGCGGATG TCGGTGTGAG CGAGCGGCGC
51 CTGAACACAC GGGGGCTGCC GAGCGCCTGA CCCGGGCTCG CGCCAGAGCC
101 TGCAACGAGC TCCGGGGCCC CACACCCGCT ACGGTGGCCC TGCGCCCGTT
151 GCTACTGAGG CGGCGTGCTC TGCATTCTTC GCTGTCCAGG CCTGCGCGCT
201 CTGGTGTCTG CTGGCTCCTC CTTGCTCGCC TGCTCCCTCC TGCTTGCCCTG
251 AGTCACCGCC GCCGCCGCCG CCACAGCCAT GGCCGAGAGT GGTGAAAGCG
301 GCGGTCTCTC GGGCTCCAG GATAGCGCGG CCGGAGCCGA AGGTGCTGGC
351 GCGGCGCGCG CCGCTGCCTC CGCGGAGCCC AAAATCATGA AAGTCAACCGT
401 GAAGACCCCG AAGGAAAAGG AGGAATTCGC CGTGCCCGAG AATAGCTCCG
451 TCCAGCAGTT TAAGGAAGAA ATCTCTAAAC GTTTTAAATC ACATACTGAC
501 CAACTTGTGT TGATATTTGC TGGAAAAATT TTGAAAGATC AAGTACCTTT
551 GAGTCAGCAT GGAATTCATG ATGGACTTAC TGTTCACTTT GTCATTAAAA
601 CACAAAACAG GCCTCAGGAT CATTGAGCTC AGCAAACAAA TACAGCTGGA
651 GGCAATGTGA CTACATCATC AACTCCTAAT AGTAACTCTA CATCTGGTTC
701 TGCTACTAGC AACCCCTTTG GTTTAGGTGG CCTTGGGGGA CTTGCAGGTC
751 TGAGTAGCTT GGGTTTGAAT ACTACCAACT TCTCTGAATC ACAGAGTCAG
801 ATGCAGCGAC AACTTTTGTC TAACCTGAA ATGATGGTCC AGATCATGGA
851 AAATCCCTTT GTTCAGAGCA TGCTCTCAAA TCCTGACCTG ATGAGACAGT
901 TAATTATGGC CAATCCACAA ATGCAGCAGT TGATACAGAG AAATCCAGAA
951 ATTAGTCATA TGTTGAATAA TCCAGATATA ATGAGACAAA CGTTGGAAC
1001 TGCCAGGAAT CCAGCAATGA TGCAGGAGAT GATGAGGAAC CAGGACCCGAG
1051 CTTTGAGCAA CCTAGAAAGC ATCCCAGGGG GATATAATGC TTTAAGGCGC
1101 ATGTACACAG ATATTCAAGG ACCAATGCTG AGTGTGTCAC AAGAGCAGTT
1151 TGGTGGTAAAT CCATTGCTT CTTTGGTGAG CAATACATCC TCTGGTGAAG
1201 GTAGTCAACC TTCCCGTACA GAAATAGAG ATCCACTACC CAATCCATGG
1251 GCTCCACAGA CTTCCACAGG TATCATCAGT TCCAGCGGCA CTGCCAGCAC
1301 TGTGGGTGGC ACTACTGGTA GTACTGCCAG TGGCACTTCT GGGCAGAGTA
1351 CTACTGCCCC AAATTGGTG CTTGGAGTAG GAGCTAGTAT GTTCAACACA
1401 CCAGGAATGC AGAGCTTGTT GCAACAATAA ACTGAAAACC CACAACATGAT
1451 TGTAAACATG TTGTCTGCCC CCTACATGAG AAGCATGATG CAGTCACTAA
1501 GCCAGAATCC TGACCTTGCT GCACAGATGA TGCTGAATAA TCCCTTATTT
1551 GCTGGAATC CTCAGCTTCA AGAACAATG AGACAACAGC TCCCACTTT
1601 CCTCCAACAA ATGCAGAATC CTGATACACT ATCAGCAATG TCAAAACCTTA
1651 GAGCAATGCA GGCCTTGTTA CAGATTCAGC AGGGTTTACA GACATTAGCA
1701 ACGGAAGCCC CGGGCCTCAT CCCAGGGTTT ACTCCTGGCT TGGGGGCATT
1751 AGGAAGCACT GGAGGCTCTT CGGGAACATA TGGATCTAAC GCCACACCTA
1801 GTGAAAACAC AAGTCCACA GCAGGAACCA CTGAACCTGG ACATCAGCAG
1851 TTTATTCAGC AGATGCTGCA GGCTCTTGCT GGAGTAAATC CTCAGCTACA
1901 GAATCCAGAA GTCAGATTTC AGCAACAAC TGAACAACCTC AGTGCAATGG
1951 GATTTTGAAG CCGTGAAGCA AACTTGCAAG CTCTAATAGC AACAGGAGGT
2001 GATATCAATG CAGCTATTGA AAGGTTACTG GGCTCCAGC CATCATAGCA
2051 GCATTTCTGT ATCTTGAAAA AATGTAATTT ATTTTGATA ACGGCTCTTA
2101 AACTTTAAAA TACCTGCTTT ATTTCAATTT GACTCTTGA ATTCTGTGCT
2151 GTTATAAACA AACCCAATAT GATGCATTTT AAGGTGGAGT ACAGTAAGAT
2201 GTGTGGGTTT TTCTGTATTT TTCTTTTCTG GAACAGTGGG AATTAAGGCT
2251 ACTGCATGCA TCACTTCTGC ATTTATTGTA ATTTTAAAA AACATCACCT
2301 TTTATAGTTG GGTGACCAGA TTTTGTCCCT CATCTGTCCA GTTTATTGTC
2351 TTTTAAACA TTAGCCTATG GTAGTAATTT ATGTAGAATA AAAGCATTAA
2401 AAAGAAGCAA ATCATTGCA CTCTATAATT TGTGGTACAG TATTGCTTAT
2451 TGTGACTTTG GCATGCATTT TTGCAACAAA TGCTGTAAAG TTTATCTAC
2501 TGATAATTTT GTTTTATTG TATACAATAT AGAGTATGCA CATTGCGGAC
```



```

2551 TGCATTCTTG GAAACATACT GCAATAGGCT CTCTGAGCAA AACACCTGTA
2601 ACTAAAAAAG TGAAGATAAG AAAATACTCT TAAAGCTGAG TATTCCTAA
2651 TTGTATAGAA TCTTACAGCA TCTTTGACAA ACATCTCCCA GCAAAAGTGC
2701 CGGTTAGTCA GGTTTGTTGA AAATACAGTA GAAAGCTGA TTCTGGTTAT
2751 CTCTTTAAGG ACAATTAATT GTACAGACAC ATAATGTAAC ATTGTCTCAA
2801 CATTCAATCA CAGATTGACT GTAAATTACC TTAATCTTTG TGCAGACTGA
2851 AGGAACACTG TAGTATACCC CAAAGTGCAT TTGCCTAGGA CTTCTCAGCT
2901 TCTCCCATAG GTAGTTTAAC AGGCATTAAA ATTTGTAATT GAAATGTTGC
2951 TTTCACTCAA AAAAAAATAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 279 bp to 2045 bp; peptide length: 589
 Category: similarity to known protein

```

1 MAESGESGGP PGSQDSAAGA EGAGAPAAAA SAEPKIMKVT VKTPKEKEEF
51 AVPENSSVQQ FKEEISKRFK SHTDQLVLIF AGKILKDQDT LSQHGHDGL
101 TVHLVIKTON RPQDHSAAQT NTAGGNVTTS STPNSTSTSG SATSNPFGLG
151 GLGGLAGLSS LGLNTTNFSE LOSQMQRQLL SNPEMMVQIM ENPFVQSMLS
201 NPDLMRQLIM ANPQMQLIQ RNPEISHMLN NPDIMRQTL LARNPAMMQE
251 MMRNQDRALS NLESIPGGYN ALRRMYTDIQ EPMLSAAQEQ FGGNPFASLV
301 SNTSSGEGSQ PSRTENRDPL PNPWAPQTSQ SSSASSGTAS TVGGTTGSTA
351 SGTSGGSTTA PNLVPGVGAS MFNTPGMQSL LQITENPQL MQNMLSAPYM
401 RSMQSLSON PDLAAQMMLN NPLFAGNPQL QEOMRQQLPT FLOQMNPDT
451 LSAMSNPRAM QALLQIQGL QTLATEAPGL IPGFTPLGA LGSTGGSSGT
501 NGSNATPSEN TSPTAGTTEP GHQQFIQQL QALAGVNPQL QNPEVRFQQQ
551 LEQLSAMGFL NREANLQALI ATGGDINAAI ERLGSSQPS

```

BLASTP hits

Entry CE1_1 from database TREMBL:
 "F15C11.2"; *Caenorhabditis elegans* cosmid VF15C11L
 Length = 293
 Score = 454 (159.8 bits), Expect = 4.4e-43, P = 4.4e-43
 Identities = 81/162 (50%), Positives = 113/162 (69%)

Entry S54583 from database PIR:
 ubiquitin-like protein DSK2 - yeast (*Saccharomyces cerevisiae*)
 Length = 373
 Score = 278 (97.9 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 100/307 (32%), Positives = 155/307 (50%)

Entry AB015344_1 from database TREMBLNEW:
 gene: "HRIHFB2157"; *Homo sapiens* HRIHFB2157 mRNA, partial cds.
 Score = 1135, P = 3.6e-115, identities = 227/301, positives = 253/301

Alert BLASTP hits for DKFZphfbr2_312, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_312, frame 3

Report for DKFZphfbr2_312.3

```

[LENGTH]      589
[MW]           62489.22
[pI]           5.02
[HOMOL]        TREMBL:AB015344_1 gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial
cds. 1e-121
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YMR276w] 2e-17

```



```

(FUNCAT)      30.10 nuclear organization      [S. cerevisiae, YMR276w] 2e-17
(BLOCKS)      BL00299 Ubiquitin family proteins
(SUPFAM)      unassigned ubiquitin-related proteins 5e-16
(SUPFAM)      ubiquitin homology 5e-16
(PROSITE)     MYRISTYL 24
(PROSITE)     CK2_PHOSPHO_SITE 9
(PROSITE)     GLYCOSAMINOGLYCAN 1
(PROSITE)     PKC_PHOSPHO_SITE 3
(PROSITE)     ASN_GLYCOSYLATION 7
(PFAM)        Ubiquitin family
(KW)          Irregular
(KW)          3D
(KW)          LOW_COMPLEXITY 23.43 %

```

```

SEQ      MAESGESGGPPGSQDSAAGAEGAGAPAAAAASAEPKIMKVTVKTPKEKEEFAVPENSSVQQ
SEG      ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
laaraA   .....CEEEEEETTTCEEEECTTTTBHHH

```

```

SEQ      FKEEISKRFKSHTDQLVLI FAGKILKDQDTLSQHGIHDGLTVHLVIKTQNRPDHSAQQT
SEG      .....
laaraA   HHHHHHHHCCCGGEEEEETTECTTTTBGGGGCCTTTTEEEEBEBC.....

```

```

SEQ      NTAGGNVTTSSTPNSNSTSGSATSNPFGGLGGLAGLSSGLNTTNFSELQSQMQRQLL
SEG      .....
laaraA   .....

```

```

SEQ      SNPEMMVQIMENPFVQSMLSNPDLMRQLIMANPQMQLIQRNPEISHMLNPNPDIMRQTL
SEG      .....
laaraA   .....

```

```

SEQ      LARNPAMQEMMRNQDRALSNLESIPGGYNALRRMYTDIQEPMLSAAQEQFGGNPFASLV
SEG      .....
laaraA   .....

```

```

SEQ      SNTSSGEGSQPSRTENRDPLPNPWAPQTSQSSSASGTASTVGGTTGSTASGTSGQSTTA
SEG      .....
laaraA   .....

```

```

SEQ      PNLVPGVGASMFNTPGMQSLQITENPQLMQNMLSAPYMRSMQSLSQNPDLAAQMMLN
SEG      .....
laaraA   .....

```

```

SEQ      NPLFAGNPQLQEQMRQLPTFLQMQNPDTLSAMSNPRAMQALLQIQQGLQTLATEAPGL
SEG      .....
laaraA   .....

```

```

SEQ      IPGFTPGLGALGSTGGSGTNGSNATPSENTSPTAGTTEPGHQQFIQQLQALAGVNPQL
SEG      .....
laaraA   .....

```

```

SEQ      QNPEVRFQQLEQLSAMGFLNREANLQALQIATGGDINAIAERLLGSQPS
SEG      .....
laaraA   .....

```

Prosite for DKFzphfbr2_312.3

PS00001	55->59	ASN_GLYCOSYLATION	PDOC00001
PS00001	126->130	ASN_GLYCOSYLATION	PDOC00001
PS00001	136->140	ASN_GLYCOSYLATION	PDOC00001
PS00001	164->168	ASN_GLYCOSYLATION	PDOC00001
PS00001	167->171	ASN_GLYCOSYLATION	PDOC00001
PS00001	302->306	ASN_GLYCOSYLATION	PDOC00001
PS00001	501->505	ASN_GLYCOSYLATION	PDOC00001
PS00002	305->309	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00005	43->46	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	71->75	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	200->204	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	572->576	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	12->18	MYRISTYL	PDOC00008

PS00008	19->25	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	140->146	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	162->168	MYRISTYL	PDOC00008
PS00008	267->273	MYRISTYL	PDOC00008
PS00008	293->299	MYRISTYL	PDOC00008
PS00008	308->314	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	492->498	MYRISTYL	PDOC00008
PS00008	495->501	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008

Pfam for DKFZphfbr2_312.3

HMM_NAME Ubiquitin family

HMM *MQIFVKTLtGRTcTFEVepQEtVeqIKQHieekEGIPPeQQRLLIFaGRQ
M ++VKT + +F V+++ V Q+K+ I+ +Q +LIFAG+

Query 37 MKVTVKTPK-EKEEFVAVPENSSVQQFKEEISKRFKSHTDQLVLIFAGKI 84

HMM LEDeKTLsDYNiggeSTLHLVlR*
L D TLS+++I + T+HLV++

Query 85 LKDQDTLSQHGIHDGLTVHLVIK 107

DKF2phfbr2_62b11

group: signal transduction

DKF2phfbr2_62b11 encodes a novel 655 amino acid putative GTPase-activating protein, related to human chimaerins.

The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to CHIMAERIN

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="4"

Insert length: 4593 bp

Poly A stretch at pos. 4571, polyadenylation signal at pos. 4553

```
1 GGGGGAGTTT GAAGACAGAA AGGAAAGGGG AGAAACCTGC AGAGAGCATC
51 AAAGGATGGG GGGTGCTATA AAAGAAGCAG GGGGGTCCTT TGAAAGAAAT
101 CTATCATGCA CTGAAATGCT TTCTGGAGAA GGTGCCGTTA TTTTCTCTCC
151 CTCTTGCTCA GATGAAAGGA GCCAGCAAGG ACAGTCCTGA AATATTCCTC
201 AGGGGACTTT TTGTCATTGT TCCTCTTTCC TCTTGACACAG AGCTATTTGC
251 TGACCTTTCC AGAGGAATCT CAGTCCAGCT GAGAAGACAG TTCTTAATAA
301 AAACAAAAAA ATGCAAAAAC CAATTCCTGC TGTTTGAATG GGAATGGTAG
351 CTTGCTTGCT GCAGTTCTTT TCCTGTGACA TTTTGGAAATG TCTGCAGAAA
401 CTTAAAAAAA AGAAAAAAA AACCTTAAAA ACTCCCTGGA TTAGGCAAGA
451 GAAAGAGGAG TTTTTTTTGT CTAACACAGG GTAAATGAGA GGTGGTAACT
501 TATCCCTAAG CCAGGACCTG GATGATCAAA ACCTTCAAA TCTAGGGATC
551 AGCACTTCAA AAATAACAAG TAAACAAGCA TGAGGAGTGG CTGTTGGGTT
601 TCGCTCAGAG GCAGGTTTTA AAGGAAGCCA AAACCGGGTT CAGAACTTCA
651 GGCCTGTACG ATGCCTGAAG ACCGGAATTC TGGGGGGTGC CCGGCTGGTG
701 CCTTAGCCTC AACTCCTTTC ATCCCTAAAA CTACATACAG AAGAATCAAA
751 CGGTGTTTTA GTTTTCGGAA AGGCATTTTT GGACAGAAAC TGGAGGATAC
801 TGTTCGTTAT GAGAAGAGAT ATGGGAACCG TCTGGCTCCG ATGTTGGTGG
851 AGCAGTGGGT GGACTTTATC CGACAAAGGG GGCTGAAAGA AGAGGGTCTC
901 TTTGCAGTGC CAGGCCAGGC TAATCTTGTT AAGGAGCTCC AAGATGCCTT
951 TGACTGTGGG GAGAAGCCAT CATTTGACAG CAACACAGAT GTACACACGG
1001 TGGCATCACT TCTTAAGCTG TACCTCCGAG AACTTCCAGA ACCAGTTATT
1051 CCTATATGCGA AGTATGAAGA TTTTGTGCA TGTGCCAAAC TGCTCAGCAA
1101 GGAAGAGGAA GCAGGTGTTA AGGAATTAGC AAAGCAGGTG AAGAGTTTGC
1151 CAGTGGTAAA TTACAACCTC CTCAAGTATA TTTGCAGATT CTTGGATGAA
1201 GTACAGTCTC ACTCGGGAGT TAACAAAATG AGTGTGCAGA ACTTGGCAAC
1251 GGTCTTTGGT CCTAATATCC TGCGCCCAA AGTGGGAAGAT CCTTTGACTA
1301 TCATGGAGGG CACTGTGGTG GTCCAGCAGT TGATGTCAGT GATGATTAGC
1351 AAACATGATT GCCTCTTTCC CAAAGATGCA GAACTACAAA GCAAGCCCCA
1401 AGATGGAGTG AGCAACAACA ATGAAATTCA GAAGAAAGCC ACCATGGGGC
1451 TGTTACAGAA CAAGGAGAAC AATAACACCA AGGACAGCCC TAGTAGGCAG
1501 TGCTCCTGGG ACAAGTCTGA GTCACCCAG AGAAGCAGCA TGAACAATGG
1551 ATCCCCCACA GCTCTATCAG GCAGCAAAAC CAACAGCCCA AAGAACAGTG
1601 TTCACAAGCT AGATGTGTCT AGAAGCCCCC CTCTCATGGT CAAAAGAAGC
1651 CCAGCCTTTA ATAAGGGTAG TGGGATAGTT ACCAATGGGT CCTTCAGCAG
1701 CAGTAATGCA GAAGGTCTTG AGAAAACCCA AACCACCCCC AATGGGAGCC
1751 TACAGGCCAG AAGGAGCTCT TCACTGAAGG TATCTGGTAC CAAAATGGGC
1801 ACGCACAGTG TACAGAATGG AACCGTGCGC ATGGGCATTT TGAACGCGA
1851 CACACTCGGG AACCCACAAA ATGTTGAAA CATGAGCTGG CTGCCAAATG
1901 GCTATGTGAC CCTGAGGGAT AACAAAGCAGA AAGAACAAGC TGGAGAGTTA
1951 GGCCAGCACA ACAGACTGTC CACCTATGAT AATGTCCATC AACAGTTCTC
2001 CATGATGAAC CTGTATGACA AGCAGAGCAT TGACAGTGCT ACCTGGTCCA
2051 CTTCTCTCTG TGAATCTCC CTCCCTGAGA ACTCCAACCT CTGTCGCTCT
2101 TCTACCACCA CCTGCCAGA GCAAGACTTT TTTGGGGGGA ACTTTGAGGA
2151 CCCTGTTTTG GATGGGCCCC CGCAGGACGA CCTTCCCAC CCCAGGACT
2201 ATGAAAGCAA AAGTGACCA AGGAGTGTGG GAGGTGGAAG TAGTCGTGCC
2251 ACCAGTAGCA GTGACAACAG TGAGACATTT GTGGGCAACA GCAGCAGCAA
2301 CCACAGTGCA CTGCACAGTT TAGTTTCCAG CCTGAAACAG GAAATGACCA
2351 AACAGAAGAT AGAGTATGAG TCCAGGATAA AGAGCTTAGA ACAGCGAAAC
2401 TTGACTTTGG AAACAGAAAT GATGAGCCTC CATGATGAAC TGGATCAGGA
2451 GAGGAAAAAG TTCACAATGA TAGAAATAAA AATGCGAAAT GCCGAGCGAG
2501 CAAAAGAAGA TGCCGAGAAA AGAAATGACA TGCTACAGAA AGAAATGGAG
2551 CAGTTTTTTT CCACGTTTGG AGAACTGACA GTGGAACCCA GGAGAACCGA
```



```

2601 GAGAGGAAAC ACAATATGGA TTCAGTGAGC CTGCTTTCGC CTGCTGTCTC
2651 TGATGGCTCT GGCAAGGACT CCAGGGATTG TGGTGGGATA TGACTTAGAA
2701 CCAGGTGGCT GGTACCTGG ATGTACAGAA GTCTAACTGG TGAAGGAATA
2751 TCATTTACAG ACATTAACA TCCATATCTG CAATGTGTAC CAAAGTTATA
2801 TCATGCCCCA TAATGCTACT GTCAAGTGT ACAACTGGAT ATGTGTATAT
2851 AGAGTAGTGT TTCAAAAGTA AACTAAAAAT GAGAAGCATA TTTCAAGAAT
2901 TATTTTATTG CAAGTCTTGT ATTTAAATGT TAAATCAATA TGTGTGTGCA
2951 ATTTAGCTTG CTTTCAAGCT TCACCCCTTG CACTTAACAT AAGCTATTTT
3001 TGGCATTGTG TTATCATCGG CTTATTTTAT AGATCAATAT TTTTATTTC
3051 CTTTTTTGCT GAGGAAATGA AGATAAGCAA AAATATAAAT ATATATATAA
3101 ATATATAGAG TATTTAAACC AGAAGAATAC TTTGTGGCTG TGCTGTTGT
3151 GCCAATAGAT TTTGTCATGA CCAAAAAGAG AAATGTAAAT AGTTTTATAA
3201 AATACAGTCG AATCACCAGG AACCTTTGAG CTGCTTTTAA AATTCTTCCC
3251 CTGGCACCAC TCAGTTTTGC TTTTGCGAGG CGATTTGACA TAGGAACTTT
3301 GAGACTCCAT GAGAAAGTCC CTTTCTGAGG CCCACTGTCT ACCTTGCCAG
3351 ATCCTCAGTG CGTATCGCCA ATGCAGGATG CTCCTTAGAA AAGAAAAAAT
3401 GGTAAGGATG GGCATTTAAC GATTCAGGCT TTGAATTACT CTGCTCCCTC
3451 GGACCGAATC TCTTTAACTG CTGGATAGTT TTAGAGGAAT TCTCTGCTA
3501 CTTAGGTACT GGGAAACAAT GCTTGCTAAA CCATGCCAC GTGAGCACCT
3551 GTCTCCCACT CAAACCTCTC CCATCTCCCA ACAACTGCAC TTTAGAATAC
3601 CAGCAGTGAA ATGGTATTAC TGTTCCTC TGAGTGAAC TGCTAGAGTA
3651 TATGTCACGT AGTGACATT TTTTCTCACT CAGGCTATTG CCATCTGGGA
3701 TTCTCTCCCT ACTACAGCTG GCAAAGTTGG TTTGAGCAA GAAGATAGTG
3751 GGAGGGGGCC AGGCTGCAGG AGAAGGAGAA AAGTTTAGAA GAAACAAACC
3801 ATTTTGCTTC TAAATTTGAC AGTATCACTT TCCTGTAAA ACATACAATA
3851 ATTTTAAAG GTGAATGCCT AAAGTTCCAA TTTTAGCAA TATGGGAACC
3901 TCAGCAATGC TAATTTCTA GAAAAACCCA GGGCTCTTTG GAGCTAGAGT
3951 TTTGGGAGAA CAGTCTTCA CAATAAGGCA ATGGTTTGA GAGGCCAGGC
4001 AAATAATCTT TCTCACCCTA GAACAAAAGG TTACAAAAGG CATAATCGGA
4051 AATAGAGACT ACATACTTGA GTTTATGGGG TTTGTGTTGT TTGAAGGTTT
4101 AATGCTTGCA TGTGTTTAT TATTTTCAAG AGGGAAGTG GTCTGTACTG
4151 CTTTCATCCT TGCCACTGTC TTTTACTCT CCCACTGAGC
4201 AAGCGTCTGT GGTCTATGG TATCAACCAG TATCTTTATA GCAATAATTT
4251 CTTTAAATCC CTTTCTCTC TCTTCCAAT TATTTAACCA GTTACTTCCA
4301 CCTGGACATA CGATAGGAAA TTCAAACCTCA AAATATGAAA ATTGATCTTA
4351 ATAACCTCTC CTTTATATCT TTTCACTAT TTCCAGTCT TATCATAGTT
4401 GATAAAACCC TCAGACTCAT CCAGAAAGCT ATATGATGCA CTAGTAAAAA
4451 AAACAAAGAT ATTTAAACTG CTTGGGTTCA AATGGTATAC AATTGCCAG
4501 CTGTTACTGA ACCTTCTATG CATAACTTTT TTTTCTCTCT GTGCAATTGG
4551 AATAATAAAA ATACTACTCC CATAAAAAAA AAAAAAAA AAC

```

BLAST Results

Entry G38474 from database EMBLNEW:
 SHGC-58303 Human Homo sapiens STS genomic, sequence tagged site.
 Score = 2175, P = 1.2e-92, identities = 439/441

Medline entries

97476250:
 Beta2-chimaerin is a high affinity receptor for the phorbol ester tumor promoters.

Peptide information for frame 1

ORF from 661 bp to 2625 bp; peptide length: 655
 Category: similarity to known protein

```

1 MPEDRNSGGC PAGALASTPF IPKTTYRRIK RCFSFRKGF GQKLEDTVRY
51 EKRYGNRLAP MLVEQCVDFI RQRGLKEEGL FRLPGQANLV KELQDAFDCG
101 EKPSFDSNTD VHTVASLLKL YLRELPEPVI PYAKYEDFLS CAKLLSKEEE
151 AGVKELAKQV KSLPVVNYNL LKYICRFLDE VQSYSGVNKM SVQNLATVFG
201 PNILRPKVED PLTIMEGTVV VQQLMSVMIS KHDCLFPKDA ELQSKPQDGV
251 SNNNEIQKKA TMGLLQNKEN NNTKDSPSRQ CSWDKSESPQ RSSMNNGSPT
301 ALSGSKTNSP KNSVHKLDVS RSPPLMVKKN PAFNKGSGIV TNGSFSSSNA
351 EGLEKTQTPP NGSLQARRSS SLKVSQTKMG THSVQNGTVR MGILNSDTLG
401 NPTNVRNMSW LPNGYVTLRD NKQEQAGEL GQHNRLSTYD NVHQQFSMMN
451 LDDKQSIDSA TWSTSSCEIS LPENSNSCRS STTTCPEQDF FGGNFEDPVL
501 DGPPQDDLSH PRDYESKSDH RSVGGRSSRA TSSSDNSETF VGNSSSNHSA
551 LHSLVSSLKQ EMTRQKIEYE SRIKSLEQRN LTLETEMMSL HDELDQERKK

```


601 FTMIEIKMRN AERAKEDAEK RNDMLQKEME QFFSTFGELT VEPRTTERGN
651 TIWIQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_62b11, frame 1

SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053., N = 3, Score = 661, P = 2.4e-89

TREMBL:HSU90908_1 product: "unknown"; Human clones 23549 and 23762 mRNA, complete cds., N = 1, Score = 348, P = 1.1e-29

PIR:S29128 N-chimerin - rat, N = 1, Score = 286, P = 2.8e-24

PIR:S29956 beta-chimerin - rat, N = 1, Score = 279, P = 1.6e-23

TREMBL:AB014572_1 gene: "KIAA0672"; product: "KIAA0672 protein"; Homo sapiens mRNA for KIAA0672 protein, complete cds., N = 1, Score = 314, P = 1e-24

>SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053.
Length = 638

HSPs:

Score = 661 (99.2 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 122/209 (58%), Positives = 160/209 (76%)

Query: 38 GIFGQKLEDTVRYEKRYGNRLAPMLVEQCVDVFIRQRLKEEGLFRLPGQANLVKELQDAF 97
G+FGQ+L++TV YE+++G L P+LVE+C +FI + G EEG+FRLPQG NLVK+L+DAF
Sbjct: 148 GVFGQRLDETVAEYEQKFGPHLVPIVLEKCAEFIEHGRNEEGIFRLPGQDNLVKQLRDAF 207

Query: 98 DCGEKPSFDSNTDVHTVASLLKLYLRELPEPVIPIYAKYEDFLSCAKLLSKEEEAGVKELA 157
D GE+PSFD +TDVHTVASLLKLYLR+LPEPV+P+++YE FL C +L + +E +EL
Sbjct: 208 DAGERPSFDRDNTDVHTVASLLKLYLRDLPEPVVPSQYEGFLCGQLTNADEAKAQQELM 267

Query: 158 KQVKSIPVVYNLLKYICRFLDEVQSYGVNKMVSQNLATVFGPNILRPKVEDPLTIMEG 217
KQ+ LP NY+LL YICRFL E+Q VNKMSV NLATV G N++R KVEDP IM G
Sbjct: 268 KQLSILPRDNYSLSYICRFLHEIQLNCVANKMSVDNLATVIGVNLIRSKVEDPAVIMRG 327

Query: 218 TVVVQQLMSVMISKHDCLEPKDAELQSKP 246
T +Q++M++MI H+ LFPK ++ P
Sbjct: 328 TPQIQRVMTMMIRDHEVLFPKSKDIPLSP 356

Score = 210 (31.5 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 45/115 (39%), Positives = 73/115 (63%)

Query: 531 TSSSDNSETFVGNSSSNHSLHSL---VSSLKQEMTKQKIEYESRIKSLEQRNLTLTETEM 587
T +S NSET G +S + SL V L++E+ QK YE +IK+LE+ N + ++
Sbjct: 523 TLASPNSETGPGKNSGEEIIDLQRMVQELRKEIETQKQMYEEQIKNLEKENYDVWAKV 582

Query: 588 MSLHDELQERKKFTMIEIKMRNAERAKEDAEKRNNDMLQKEMEQQFFSTFGELTVE 642
+ L++EL++E+KK +EI +RN ER++ED EKRN L++E+++F + E E
Sbjct: 583 VRLNEELEKEKKKSAALEISLRNMERSREDVEKRNKALEEEVKEFVKSMKEPKTE 637

Score = 70 (10.5 bits), Expect = 1.2e-74, Sum P(3) = 1.2e-74
Identities = 28/121 (23%), Positives = 54/121 (44%)

Query: 528 SRATSSSDNSETFVGNSSSNHSLVSSLKQE-MTKQKIEYESRIKSLEQRNL-TLET 585
S+ TS+ DN + G+ SAL S K + + E K+ + + +L+
Sbjct: 489 SQRTSTYDNVPSLPSPGEEASALSSQACDSKGDTLASPNSETGPGKNSGEEIIDLQ 548

Query: 586 EMMSLHDELQERKKFTMIEIKMRNAERAKEDAEKRNNDMLQKEMEQQFFSTFGELTVEPRR 645
+ L E++ +++ M E +++N E+ D + L +E+E+ L + R
Sbjct: 549 MVQELRKEIETQKQ---MYEEQIKNLEKENYDVWAKVVRLNEELEKEKKKSAALEISLRN 605

Query: 646 TER 648
ER
Sbjct: 606 MER 608

Score = 53 (8.0 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 31/111 (27%), Positives = 46/111 (41%)

Query: 344 SFSSSNAEGLEKTQTPNGSLQARRSSSLKVSQTKMGTHSVQNG---TV--RMGILNSD 397
SFSS ++ + T T A S KV K G +Q+ T+ R L S
Sbjct: 388 SFSSMTSDS-DTTSPTGQQPSDAFPEDSSKVPREKPGDWKMQSRKRTQTLNRRKCFITSA 446

Score = 53 (8.0 bits), Expect = 3.5e-14, Sum P(3) = 3.5e-14
Identities = 32/125 (25%), Positives = 56/125 (44%)

Query: 354 EKTQTTPN 361
++TQT PN
Sbjct: 431 KRTQTLPN 438

Report for DKFZphfbr2 62b11.1

```

SEQ      MPEDRNSGGCPAGALASTPFIPTKTTYRRIKRCFSFRKGI FGQKLEDTVRYEKRYGNRLAP
SEG      .....
COILS    .....
1rgp-    .....C

SEQ      MLVEQCVDFIQRGLKEEGLFRLPGQANLVKELQDAFDCGEKPSFDSNTDVHTVASLLKL
SEG      .....
COILS    .....
1rgp-    HHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHHHCCCCGGGGCCCCHHHHHHHHHH

SEQ      YLRELPEPVIPYAKYEDFLSCKALLSKEEEAGVKELAKQVKS LPVVNYNLLKYICRFLE
SEG      .....

```



```

COILS .....
lrgp- HHHHTTTTTTGGGHHHHH---TTTCGGGHHHHHHHHHCCCHHHHHHHHHHHHHHH
SEQ    VQSYSGVNKMSVQNLATVFGPNILRPKVEDPLTIMEGTVVVQQLMSVMISKHDCLFKDA
SEG    .....
COILS .....
lrgp- HHHHHHHHCCCHHHHHHHGGGCC.....
SEQ    ELQSKPDQGVSNNEIQKKATMGLLQNKENNTKDSPSRQCSWDKSESQRSSMNGSPT
SEG    .....
COILS .....
lrgp- .....
SEQ    ALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAFNKSGIVTNGSFSSSNAEGLEKTQTP
SEG    .....
COILS .....
lrgp- .....
SEQ    NGSQARRSSSLKVSQTKMGTHSVQNGTVRMGILNSDTLGNPTNVRNMSWLPNGYVTLRD
SEG    .....
COILS .....
lrgp- .....
SEQ    NKQKEQAGELGQHNLSTYDNVHQQFSMMNLDDKQSIDSATWSTSSCEISLPENSNSCRS
SEG    .....xxxxxxx
COILS .....
lrgp- .....
SEQ    STTTCPEQDFGGNFEDPVLGPPQDDLSPRDYESKSDHRSVGGRRSSRATSSSDNSETF
SEG    xxxxxx.....xxxxxxxxxxxxxxxxxxxxx...
COILS .....
lrgp- .....
SEQ    VGNSSSNHSALHSLVSSSLQEMTKQKIEYESRIKSLEQRNLTLETMMSLHDELQERKK
SEG    ..xxxxxxxxxxxxxxxxxxxxx.....
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
lrgp- .....
SEQ    FTMIEIKMRNAERAKEDAEKRNMLQKEMEQQFSTFGELTVEPRTERGNTIWIQ
SEG    .....
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
lrgp- .....

```

Prosite for DKFZphfbr2_62b11.1

PS00001	271->275	ASN_GLYCOSYLATION	PDOC00001
PS00001	342->346	ASN_GLYCOSYLATION	PDOC00001
PS00001	361->365	ASN_GLYCOSYLATION	PDOC00001
PS00001	386->390	ASN_GLYCOSYLATION	PDOC00001
PS00001	407->411	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	547->551	ASN_GLYCOSYLATION	PDOC00001
PS00001	580->584	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	367->371	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	599->603	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	47->50	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	388->391	PKC_PHOSPHO_SITE	PDOC00005
PS00005	417->420	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	527->530	PKC_PHOSPHO_SITE	PDOC00005
PS00005	557->560	PKC_PHOSPHO_SITE	PDOC00005
PS00005	646->649	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	213->217	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	348->352	CK2_PHOSPHO_SITE	PDOC00006
PS00006	417->421	CK2_PHOSPHO_SITE	PDOC00006
PS00006	437->441	CK2_PHOSPHO_SITE	PDOC00006
PS00006	465->469	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	516->520	CK2_PHOSPHO_SITE	PDOC00006
PS00006	532->536	CK2_PHOSPHO_SITE	PDOC00006

PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	635->639	CK2_PHOSPHO_SITE	PDOC00006
PS00007	43->51	TYR_PHOSPHO_SITE	PDOC00007
PS00007	176->185	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	263->269	MYRISTYL	PDOC00008
PS00008	297->303	MYRISTYL	PDOC00008
PS00008	304->310	MYRISTYL	PDOC00008
PS00008	338->344	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	376->382	MYRISTYL	PDOC00008
PS00008	392->398	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	542->548	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_62b11.1)

DKFZphfbr2_62f10

group: intracellular transport and trafficking

DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with strong similarity to mammalian zinc transporter proteins.

The novel proteins is a membrane protein, which should be involved in the transport of Zinc across the cell membrane.

The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide.

The new protein can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation.

strong similarity to zinc transporter proteins ;
membrane regions: 5

Summary DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with similarity to zinc transporter protein.
The new protein can find clinical application in modulating Zn²⁺ uptake.

strong similarity to zinc transporter proteins

complete cDNA, complete cds, few EST hits

Sequenced by LMU

Locus: unknown

Insert length: 5422 bp

Poly A stretch at pos. 5397, polyadenylation signal at pos. 5381

```
1  GTCTAACTTT  GGAATATCA  CCTCATGCT  GTCTTCCAG  GATGCTCTC
51  TCCCTAAGTA  AGGGATGTTA  CTTCTGGAG  GGAATGCAGT  GTTGGGAATC
101 TGAAGACCCA  GCTTTGAGCT  GAATTTGCTT  TGTGATACCT  GGAGAGAAGA
151 CGTGTTTTCT  TGACAACAGC  ACAGTACCTA  GTGAGTTCAA  CAACAACGAC
201 AACAAACAGC  GCAGCTCATC  CTGGCCGTCA  TGGAGTTTCT  TGAAGAGCG
251 TATCTTGTGA  ATGATAAAGC  TGCCAAGATG  TATGCTTTCA  CACTAGAAAG
301 AAGGAGCTGC  AATGAACAC  TTCATAGCAA  TGTGGAACCT  CAACAGAAAC
351 CGGTGAATAA  AGATCAGTGT  CCCAGAGAGA  GACCAGAGGA  GCTGGAGTCA
401 GGAGGCATGT  ACCACTGCCA  CAGTGGCTCC  AAGCCACAG  AAAAGGGGGC
451 GAATGAGTAC  GCCTATGCCA  AGTGGAAACT  CTGTTCTGCT  TCAGCAATAT
501 GCTTCATTTT  CATGATTGCA  GAGGTCGTGG  GTGGGCACAT  TGCTGGGAGT
551 CTTGCTGTTG  TCACAGATGC  TGCCACCTC  TTAATTGACC  TGACCAGTTT
601 CTTGCTCAGT  CTCTTCTCCC  TGTGGTTGTC  ATCGAAGCCT  CCCTCTAAGC
651 GGCTGACATT  TGGATGGCAC  CGAGCAGAGA  TCCTTGTTGC  CCTGCTCTCC
701 ATCTCTGTGA  TCTGGGTGGT  GACTGGCGTG  CTAGTGTAAC  TGGCATGTGA
751 GCGCCTGCTG  TATCCTGATT  ACCAGATCCA  GCGGACTGTG  ATGATCATCG
801 TTTCCAGCTG  CGCAGTGGCG  GCCAACATTG  TACTAACTGT  GGTTTTGAC
851 CAGAGATGCC  TTGGCCACAA  TCACAAGGAA  GTACAAGCCA  ATGCCAGCGT
901 CAGAGCTGCT  TTTGTGCATG  CCCCTGGAGA  TCTATTTTCA  AGTATCAGTG
951 TGCTAATTAG  TGCATTATT  ATCTACTTTA  AGCCAGAGTA  TAAATAGCC
1001 GACCCAATCT  GCACATTCAT  CTTTTCATC  CTGGTCTTGG  CCAGCACCAT
1051 CACTATCTTA  AAGGACTTCT  CCATCTTACT  CATGGAAGGT  GTGCCAAAGA
1101 GCCTGAATTA  CAGTGGTGTG  AAAGAGCTTA  TTTTAGCAGT  CGACGGGGTG
1151 CTGTCTGTGC  ACTGCCTGCA  CATCTGGTCT  CTAACAATGA  ATCAAGTAAT
1201 TCTCTCAGCT  CATGTTGCTA  CAGCAGCCAG  CCGGGACAGC  CAAAGTGGTC
1251 GGAGAGAAAT  TGCTAAAGCC  CTTAGCAAAA  GCTTTACGAT  GCACTCACTC
1301 ACCATTGAGA  TGAATCTCC  AGTTGACCAG  GACCCGACT  GCCTTTTCTG
1351 TGAAGACCCC  TGTGACTAGC  TCAGTCACAC  CGTCAGTTTC  CCAAAATTTGA
1401 CAGGCCACCT  TCAAACATGC  TGCTATGCAA  TTTCTGCATC  ATAGAAAATA
1451 AGGAACCAAA  GGAAGAAATT  CATGTCATGG  TGCAATGCAT  ATTTTATCTA
1501 TTTATTTAGT  TCCATTCAAC  ATGAAGGAAG  AGGCACTGAG  ATCCATCAAT
1551 CAATTGGATT  ATATACTGAT  CAGTAGCTGT  GTTCAATTGC  AGGAATGTGT
1601 ATATAGATTA  TTCTGAGTG  GAGCCGAGT  AACAGCTGTT  TGTAACATC
1651 GGCAATACCA  AATTCATCTC  CCTTCCAATA  ATGCATCTTG  AGAACACATA
1701 GGTAAATTTG  AACTCAGGAA  AGTCTTACTA  GAAATCAGTG  GAAGGGACAA
1751 ATAGTCACAA  AATTTTACCA  AAACATTAGA  AACAAAAAT  AAGGAGAGCC
1801 AAGTCAGGAA  TAAAAGTGAC  TCTGTATGCT  AACGCCACAT  TAGAACTTGG
```



```

1851 TTCTCTCACC AAGCTGTAAT GTGATTTTTT TTTCTACTCT GAATTGGAAA
1901 TATGTATGAA TATACAGAGA AGTGCCTTACA ACTAATTTTT ATTTACTTGT
1951 CACATTTTGG CAATAAATCC CTCTTATTTC TAAATTTCTAA CTTGTTTATT
2001 TCAAAACTTT ATATAATCAC TGTTCAAAAG GAAATATTTT CACCTACCAG
2051 AGTGCCTTAA CACTGGCACC AGCCAAAGAA TGTGGTTGTA GAGACCCAGA
2101 AGTCTTCAAG AACAGCCGAC AAAAACATTG GAGTTGACCC CACCAAGTTG
2151 TTGCCACAGA TAATTTAGAT ATTTACCTGC AAGAAGGAAT AAAGCAGATG
2201 CAACCAATTC ATTCAGTCCA CGAGCATGAT GTGAGCACTG CTTTGTGCTA
2251 GACATTGGGC TTAGCACTGA AACTATAAAG AGGAATCAGA CGCAGCAAGT
2301 GCTTCTGTGT TCTGGTAGCA ACTCAACACT ATCTGTGGAG AGTAAACTGA
2351 AGATGTGCAG GCCAACATTG TGGAAATCCT ATGTCAGTGG GTTTGGTTTG
2401 GAACCTGGAC TTCTGCATTT TTAAGAGTTA CCCAGAGATG CTTCTAAAGA
2451 TGAGCCATAG TCTAGAAGAT TGTCAACCAC AGGAGTTTAT TGAGTGGGAC
2501 AGCTAGACAC ATACATTGGC AGTTACAATA GTATCATGAA TTGCAATGAT
2551 GTAGTGGGGT ATAAAAGGAA AGCGATGGAT ATTGCCGGAT GGGCATGGCC
2601 AGTGTAGTTT CACGTCATTG AGGTGACAGC TCTGCTGGAC TTTGAATTAC
2651 ATATGGAGGC TCTCCAGGAA GACGAAGAAG AGAAGGACAT TCTAGGCAAA
2701 AAGAAGACTA GGCACAAGGC ACACCTATGT TTGCTCTGTA GCTTTTAGTT
2751 GAAAAAGCAA AATACATGAT GCAAAGAAAC CTCTCCACGC TGTGATTTTT
2801 AAAACTACAT ACTTTTGTGA ACTTTATGGT TATGAGTATT GTAGAGAACA
2851 GGAGATAGGT CTTAGATGAT TTTTATGTTG TTGTCAGACT CTAGCAAGGT
2901 ACTAGAAACC TAGCAGGCAT TAATAATTGT TGAGGCAATG ACTCTGAGGC
2951 TATATCTGGG CCTTGTCAAT ATTTATCATT TATATTTGTA TTTTCTCTG
3001 AAATTTGAGG GCCAAGAAA CATTGACTTT GACTGAGGAG GTCACATCTG
3051 TGCCATCTCT GCAAAATCAAT CAGCACCAC GAAATAACTA CTTAGCATTG
3101 TGTCTGAGCT TCCCTGCTCA GTAGAGACAA ATATACTCAT CCCCCACCTC
3151 AGTGAGCTTG TTTAGGCAAC CAGGATTAGA GCTGCTCAGG TTCCCAACGT
3201 CTCCTGCCAC ATCGGGTTCT CAAAATGGAA AGAATGGTTT ATGCCAAATC
3251 ACTTTTCTCT TCTGAAGGAC CACTGAATGG TTTTGTTTTT CCATATTTTG
3301 CATAGGACGC CCTAAAGACT AGGTGACTTG GCAAACACAC AAGTGTTAGT
3351 ATAATTCTTT GCTTCTGCTT CTTTTTGAAA ATCATGTTTA GATTTGATT
3401 TAAGTCAGAA ATTCACGTGA TGTCAAGTAA TCATTATGGA GGGAGATTG
3451 TGTGTCAACC AAAGTAATTG TCCCATGGCC CCAGGGTATT TCTGTTGTTT
3501 CCCGAAATTT CTGCTTTTTT AGTCAGCTAG ATTGAAAAC CTGAACAGTA
3551 GATGTTTATA TGGCAAAATG CAAGACAATC TATAAGGGAG ATTTTAAGGA
3601 TTTTGAGATG AAAAAACAGA TGCTACTCAG GGGCTTTATG GACCATCCAT
3651 CAATTCTGAA GTTCTGACTC TCCCATTACC CTTTCCCTGG TGTGGTCAGA
3701 ACTCCAGGTC ACTGGAAGTT AGTGAATCA TGATGTTGAA TCTTTACTT
3751 CAAGACATTG TATCTCTCC AGCTATCAAA ACATTATGTA TCTTTTATG
3801 CTTTTTTTGT TTATTGTTAT ACTTTAAGTT CTGGGGTACA TGTGCGGAAC
3851 ATGTAGGTTT GTTACATAGG TATACATGTG CCATGGTGGT TTGCTGCAT
3901 CATCAACCTG TCATCTACAT TCTTTTATGT CTGTCTTTCA AAGCAACACT
3951 CTGTTCTTCT GAGTAGTGAA ATCAGGTCAA CTTTACCACC AGCCTCCATT
4001 TTTAATATGC TTCACCATCA TCCAGCACCT ACTTAAGATT TATCTAGGGC
4051 TCTGTGGTGA TGTAGGACC CATAAAAGAA ATTTATGCC TCCATATGTT
4101 TGGTTACAGA TGGGAAATGG GAATGTTGAA GGACATGAAA GAAAGGATGT
4151 TTACACATTA AGCATCAGTT CTGAAGCTAG ATTGCTGAG TTTGAATCTT
4201 AGCTCTTCCC TTTATTAGCT CTGTGACCTC GAGCTAGTTA CTTAAATGCT
4251 CTGATCCTCT ATTTCTGAT CAGTGAAACC TCCCTATTCA AATGTGTGAG
4301 AGTTTAATAA ATTAGGACAC TTAATAATGT TGGAGCAGTG CATAGCATGT
4351 AGTGTTCAGT ACATGTTAAA TGTGTTTTT TATTATGTAC AAACATGTGT
4401 GGGCACAGAA TTTTAAATCA TCTCAACTTT TGAGAAATTT TGAGTTATCA
4451 ACACCGTTCC CACAAGACAG TGGCAAAATT ATTGGTGAGA ATTAACAGC
4501 TGTTTCTCAG AGGAAGCAAT GGAGGCTTGC TGGGATAAAG GCATTTACTG
4551 AGAGGCTGTT ACCTAGTGAG AGTGATGAAT TAATTAATAA AGTCGAATCC
4601 CTTTCTGACT GTCTCTGAAA GCTTCCGCTT TTATCTTTGA AGAGCAGAAT
4651 TGTCAACCCA AGGACATTTA TTAATAAAAA GAACAACTGT CCAGTGCAAT
4701 GAAGGCAAGG TCATAGGTCT CCCAAGTCTT ACCCATTTCC TGTGAAATAT
4751 CAAGTTCTTG GCTTTTCTCT GTCATGTAGC CTCAACTTTC TCCGACCGGG
4801 TGCATTTCTT TCTCTGGTTT CTAATTGCC AGTGGCAAAT TTGGATCACT
4851 TACTTAATAT CTGTTAAATT TTGTGACCCA ACAGAGTCTT TTAGCACTGT
4901 GGTGTCAAAA AGAAAAACAC CTCCAGGCA TATACATTTT ATAGATTCTT
4951 GGAGAATGTT GCTCTCCAGC TCCATCCCA CCCAATGAAA TATGATCCAG
5001 AGAGTCTTGC AAAGAGACAA GCCTCATTTT CCACAATTAG CTCTAAAGTG
5051 CCTCCAGGAA ATGATTTTCT CAGCTCATCT CTCTGTATTC CCTGTTTGG
5101 ATCAGAGGGC AATCTGTTTA AATGACTAAT TACAGAAATC ATTAAGGCA
5151 CCAAGCAATG GTCATCTCTG AATACACACA TCCCAAGCTT TACAAATCCT
5201 GCCTGGCTTG ACAGTGATGA GGCCACTTAA CAGTCCAGCG CAGGCGGATG
5251 TTAATAAAAA TAAAGAGGTG ACCATCTCGG GTTTAGTTTT TTAACTTTCT
5301 GATTTACACAT TTAACGTCTG TCATTCTGTT ACTGGGCACC TGTTTAAAT
5351 CTATTTTAAA ATGTTAATGA GTGTTGTTTA AAATAAATC AGGAAAGAGA
5401 GAAAAAATAA AAAAAAATAA AC

```

BLAST Results

No BLAST result

Medline entries

97121493:

ZnT-3, a putative transporter of zinc into synaptic vesicles.

96203098:

ZnT-2, a mammalian protein that confers resistance to zinc by facilitating vesicular sequestration.

Peptide information for frame 2

ORF from 407 bp to 1366 bp; peptide length: 320
 Category: strong similarity to known protein

```

1 MYHCHSGSKP TEKGANEYAY AKWKLCSSASA ICFIFMIAEV VGGHIAGSLA
51 VVTDAAHLLI DLTSFLLSLF SLWLSSKPPS KRLTFGWHRA EILGALLSIL
101 CIWVVTGVLV YLACERLLYP DYQIQATVMI IVSSCAVAAN IVLTVVLHQR
151 CLGHNHKEVQ ANASVRAAFV HAPGDLFQSI SVLISALIIY FKPEYKIADP
201 ICTFIFISILV LASTITILKD FSILLMEGVV KSLNYSQVKE LILAVDGVLS
251 VHCLHIWSLT MNQVILSAHV ATAASRDSQV VRREIAKALS KSFTMHSILT
301 QMESPVDDQF DCLFCEDPCD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62f10, frame 2

PIR:S70632 zinc transporter ZnT-2 - rat, N = 1, Score = 884, P = 1.5e-88

TREMBL:MMU76007_1 gene: "ZnT-3"; product: "ZnT-3"; Mus musculus zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 772, P = 1.1e-76

TREMBL:HSU76010_1 gene: "ZnT-3"; product: "ZnT-3"; Human putative zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 742, P = 1.6e-73

TREMBL:MMUZNT02_1 gene: "ZnT-3"; product: "zinc transporter"; Mus musculus zinc transporter (ZnT-3) gene, complete cds., N = 1, Score = 715, P = 1.2e-70

TREMBL:CET18D3_3 gene: "T18D3.3"; Caenorhabditis elegans cosmid T18D3, N = 1, Score = 699, P = 5.9e-69

>PIR:S70632 zinc transporter ZnT-2 - rat
 Length = 359

HSPs:

Score = 884 (132.6 bits), Expect = 1.5e-88, P = 1.5e-88
 Identities = 171/326 (52%), Positives = 230/326 (70%)

```

Query:      2 YHCHSGSKPTEKGANEYAYAKWKLCSSAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61
              ++CH+      +E   A+ KL ASAIC +FMI E++GG++A SLA++TDAHLL D
Sbjct:     34 HYCHAQKDSGSHPNSEKQARRKLYVASAICLVFMIGEIIIGGYLAQSLAINTDAHLLTD 93

Query:     62 LTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 121
              S L+SLFSLW+SS+P +K + FGW RAEILGALLS+L IWVVTGVLVYLA +RL+ D
Sbjct:     94 FASMLISLFLWVSSRPATKTMNFGWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGD 153

Query:    122 YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH-----KEVQANASVRAAFVHAPG 174
              Y+I+      M+I S CAVA NI++ + LHQ  GH+H      + Q N SVRAAF+H G
Sbjct:    154 YEIKGDTMLITSGCAVAVNIIMGLALHQSGHGHSHGHSHEDSSQQQNPSVRAAFIHVVG 213

Query:    175 DLFQSI SVLISALIIYFKPEYKIADPCTFIFISILVLA LASTITILKDFSI LLMEGVPKSLN 234
              DL QS+ VL++A IYFKPEYK DPICTF+FSILVL +T+TIL+D ++LMEG PK ++
Sbjct:    214 DLLQSVGVLVAAAYIIYFKPEYKYVDPICTFILFISILVLGTTTLTILRDVILVLMEGTPKGV 273

Query:    235 YSGVKELILAVDGVLSVHCLHIWSLT MNQVILSAHVATAASRDSQVVRREIAKALS KSFT 294
              ++ VK L+L+VDGV ++H LHIW+LT+ Q +LS H+A A + D+Q V +      L F
Sbjct:    274 FTTVKNLLLSVDGVEALHSLHIWALTVAQPVLVSHIAIAQNVDAAVQLKVARDRQLQGFN 333

```


Query: 295 MHSITQMESPVDDQDCLFCEDPCD 320
 H++TIQ+ES + C C+ P +
 Sbjct: 334 FHTMTIQIESYSEDMKSCQECQGPSE 359

Pedant information for DKFZphfbr2_62f10, frame 2

Report for DKFZphfbr2_62f10.2

[LENGTH] 320
 [MW] 35053.51
 [pI] 6.48
 [HOMOL] PIR:S70632 zinc transporter ZnT-2 - rat 3e-84
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YMR243c] 2e-16
 [FUNCAT] 13.01 homeostasis of metal ions [S. cerevisiae, YMR243c] 2e-16
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YMR243c] 2e-16
 [FUNCAT] 11.07 detoxification [S. cerevisiae, YMR243c] 2e-16
 [FUNCAT] 07.04.01 metal ion transporters (cu, fe, etc.) [S. cerevisiae, YMR243c] 2e-16
 [FUNCAT] 08.04 mitochondrial transport [S. cerevisiae, YOR316c] 3e-13
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YOR316c] 3e-13
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR205w] 4e-07
 [PIRKW] transmembrane protein 2e-30
 [PIRKW] mitochondrial inner membrane 6e-12
 [PIRKW] mitochondrion 6e-12
 [PIRKW] membrane protein 1e-11
 [SUPFAM] zinc transporter ZnT-2 2e-30
 [SUPFAM] membrane protein czcD 1e-11
 [PROSITE] MYRISTYL 4
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 1
 [PROSITE] PROKAR_LIPOPROTEIN 1
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] TRANSMEMBRANE 5
 [KW] LOW_COMPLEXITY 8.12 %

SEQ MYHCHSGSKPTEKGANEYAYAKWKLCSASAICFIEMIAEVVGGHIAGSLAVVTDAHLLI
 SEGxxx
 PRD ccc
 MEMMM

SEQ DLTSFLLSLFSLWLSSKPPSKRLTFGWRAEILGALLSILCIWVVTGVLVYLACERLLYP
 SEG xxx
 PRD hhhc
 MEM MMMMMMMMMMMMM.....MM

SEQ DYQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNHKEVQANASVRAAFVHAPGDLFQSI
 SEG
 PRD ccc
 MEM MMM

SEQ SVLISALIIYFKPEYKIADPICTFIFISILVLASTITILKDFSILLMEGVKPSLNYSQVKE
 SEG
 PRD hhh
 MEM ..MM

SEQ LILAVDGVLSVHCLHIWSLTMNQVILSAHVATAASRDSQVVRREIAKALSFSFTMHSITI
 SEG
 PRD hhh
 MEMcc

SEQ QMESPVDDQDCLFCEDPCD
 SEG
 PRD eee
 MEMcc

Prosite for DKFZphfbr2_62f10.2

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00001	234->238	ASN_GLYCOSYLATION	PDOC00001
PS00004	81->85	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005

PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00007	13->21	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	228->234	MYRISTYL	PDOC00008
PS00013	125->136	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_62f10.2)

DKFZphfbr2_62n10

group: brain derived

DKFZphfbr2_62n10 encodes a novel 541 amino acid protein with similarity to Plasmodium vivax reticulocyte-binding protein 1.

The novel protein contains one Leucine Zipper, involved in protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to reticulocyte-binding protein

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="13"

Insert length: 3522 bp

Poly A stretch at pos. 3503, polyadenylation signal at pos. 3479

```
1 GGGGCGTGTT GCGGGGATTC TGAACGCTGC CATGGCTCAG ACCGTGTAGA
51 ATGTTACATT GTCGCTCACT CTGCCCATCA CGTGCCACAT TTGCTTGGGG
101 AAGGTACGTC AGCCTGTCAT ATGCATCAAC AACCATTGAT TTTGTTTCGAT
151 TTGTATTGAT TTGTGGTTGA AGAATAATAG CCAGTGTCCA GCTTGACAGAG
201 TCCCCATCAG TCCTGAAAAT CCTTGCAAAG AAATTATAGG AGGAACAAGT
251 GAAAGTGAAC CTATGCTAAG CCATACGGTC AGGAAGCATC TTCGGAAAAC
301 TAGACTTGAA TTACTACACA AAGAATATGA GGACGAAATA GATTGTTTAC
351 AGAAAGAAAGT AGAAGAGCTT AAGAGTAAAA ATCTCAGCTT GGAGTCACAG
401 ATCAAAGCTA TTCTGGATCC TTTAACCTTG GTGCAGGGCA ACCAAAATGA
451 AGACAAACAT CTAGTCACAG ATAATCCAAG TATAATTAAC CCAGAACTG
501 TAGCAGAGTG GAAGAAAAAA CTCAGAACAG CTAATGAAAT CTATGAAAAA
551 GTGAAAGATG ATGTGGATAA GCTAAAGGAG GCAAATAAAA AATTGAAATT
601 GGAAAATGGT GGTCTGGTGA GGGAGAAATT ACGACTGAAG GCTGAAAGTG
651 ATAACAGATC ACCTCAAAAG TTTGGAAGGT TTGCAGTTGC TGCTCTTCAG
701 TCCAAAGTAG AACAGTATGA GCGTGAAACC AATCGCCTCA AGAAAGCCCT
751 GGAACGAAGT GATAAGTATA TAGAGGAACT AGAATCTCAA GTTGACACAGC
801 TAAAAAATTC AAGTGAAGAG AAAGAGGCTA TGAATTCCAT TTGCCAGACA
851 GCACCTTTCTG CAGATGGCAA AGGGAGCAAA GGCAGTGAGG AGGATGTGGT
901 GTCAAAGAAT CAAGGCGATA GTGCCAGAAA GCAGCCTGGC TCATCCACCT
951 CCAGTTCTTC TCACCTAGCG AAGCCTTCCA GCAGCAGACT GTGTGACACC
1001 AGTTCCTGCA GGCAGGAAAG TACCAGCAAA GCAGACCTTA ACTGTTCTAA
1051 GAACAAAGAC CTATATCAAG AACAGGTAGA AGTAATGTTA GATGTGACAG
1101 ATACAAGTAT GGATACTTAT TTGGAAAGAG AATGGGGGAA TAAACCAAGT
1151 GACTGTGTAC CCTACAAAGA TGAAGAACTT TATGATTTTC CAGCTCCTTG
1201 TACTCCTTTG TCCCTTAGTT GCCTTCAGCT CAGTACTCCA GAAAATAGAG
1251 AGAGCTCTGT GGTCCAAGCA GGAGGTTCCA AAAAGCACTC AAACCATCTC
1301 AGAAAATTGG TGTTTGATGA TTTTGTGAT TCTTCAAATG TTTCTAATAA
1351 AGATTCTTCA GAAGATGATA TAAGTAGAAG TGAAAATGAG AAGAAATCAG
1401 AATGTTTTTC TTCCACAAAG ACAGGATTTT GGGACTGTTG TTCCACAAGC
1451 TATGCCCAAA ACTTAGATTT TGAAAGTCA GAGGGGAACA CGATAGCAAA
1501 TTCTGTTGGA GAAATATCTT CAAAATTGAG TGAGAAATCA GGCTTATGTT
1551 TATCCAAAAG GTTGAATTCT ATTGCTCTT TTGAAATGAA CCGGACAAGA
1601 ACATCCAGTG AAGCATCGAT GGATGCTGCT TACCTTGACA AAATCTCTGA
1651 GTTGGAATTC ATGATGTCAG AGTCAGACAA CAGCAAGAGC CTTTGAATA
1701 ACGGTTTTTA GTCACCTGGT TTGGATGGGT TATCAAAGTC ATCTCAAGGC
1751 AGTGAATTTT TTGAGGAACC TGATAAGTTG GAAGAAAAAA CTGAGCTAAA
1801 CCTTTCCAAA GGTTCCTCTA CTAATGATCA GTTAGAAAAT GGAAGTGAAT
1851 GGAACCCAC TTTCTTTTCT TCTCCTCTCT CCATCTGACC AAGAAATGAA
1901 TGAAGATTTT TCACTCCATT CCAGTCTTTG TCCAGTAACT AATGAATCA
1951 AACCCCAAG CTGCTTGTGT CAGACAGAGT TTTCCCAGGG CATTTTGTTA
2001 AGCAGTTCAC ATCGACTATT GGAAGATCAA AGATTGGGT CATCTTTGTT
2051 TAAGATGTCC TCAGAGATGC ACAGTCTTCA TAACCACCTT CAGTCTCCTT
2101 GGTCTACTTC CTTTGTGCCT GAAAAGAGGA ATAAAAATGT GAATCAATCA
2151 ACAAAGAGAA AAATCCAGAG CAGCCTTTCC AGTGCCAGCC CATCAAAAGC
2201 AACTAAAAGT TGACTCATT GAAAGGTGTC ATTTGTGGTT TTGCTCTGAG
2251 AGAAATAGAA AAGTTGTAA AGTTACCTTT TTTCTCATA AAAGTCTAT
2301 ACAAATTGGA ATTGATAATC TTTAGTCAAG TATCAAGTCA GGATGGTGGA
2351 TTAACCTGTA CCCAGAATAC TTATTGTTCA TTTTGAAAAG ACTTTGTTCT
2401 TTTCAATTTT ATTTGGGAGT CTTTGTGACC AGAGAAGTTA GGGAGGAGGT
2451 TATTTTTGTG TTTTGGGGTT GGTGTTGTTG TTGGTTTTGT TTTTGGTTTT
2501 GTTTTTTTAC TGAATTTGAT ATGTATCTCG GTTGATATA CATTGTTTTT
2551 TTAATAAATG TTATTTAACT GTTAGATACA GTGGCCTGTT GATAAGCCCC
2601 ACTTGCTCTC AGAAGTTGGA TTTCTTAAAT AAAACTTTTA GTGTTGTCTA
```



```

2651 TACACTGCTC AATAAGACAC TTGAGTTTAA GCTTTTCCCA GGGTGGAAAT
2701 TATTTTACCT GTCCCTTTTT ATTTATGTTT AGTGATGGCC TAGTTTTTCT
2751 GCAGGGCCAT GATGGAGAAA TAGCACTCTA GCCTTAGTCC AATATTGATT
2801 TACTTTCTTT TTTTAGGTTT TATGTATATG TTTGCATTTT TTAGCATTGT
2851 GTTTTGTCCA GTTTTGTGAA AATGTTCTGC TAGTATGAAA GAAACATTT
2901 TCTATATGAA GACATTTGTT TTATGTTAGG TAGCTTACAT TTTCTCCTCT
2951 GCGTGTGTGT GTATGTGTGT AAAATCAGAA ATTTAGCATA CTATGGAAAG
3001 AAGGCATGGA GCACTTGGGT TTAGAGGAAC CTAAACATC ATAGCTTCAT
3051 TGTTCAGAT GTAACAGGT TGAAAGAGCT CATCGCCAAG TTCCTGATCC
3101 ACTTGCATTC CAGGGGAGTT CTCTTTTGTAG TAGTATGTTT CTTGTTTGCA
3151 TGTTCCTGTT CTTTGTGGAA ACTATGCATG GTAGCATTTT TGCTTGCTGT
3201 GTTTTCCATA CTTAAGAAAA AGAGGTTTCA GTTGGCTGAT AGAATATCTT
3251 TTATGTAGGA CAAACTTTT CTGTGAAGAG TGTTGAGGGG GTGAAGATAG
3301 GTAAGAGGTA AGCACAATT TTAATTTAGG CTCTGAAAAA GTGTATTGTT
3351 CTAACAGTAT TTGGTATGCC TATATAGGTC TTTAAAAATG GGTGTTGTATG
3401 CTGTTTAATG TGCACTGAAC ATTTTACATT AATATTGTAC TGTTTACAT
3451 TAATACTGCA TGCTTTTCTA TGTGAATTGA ATAAAGAATG TCATAAGCAC
3501 TGGAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS658254 from database EMBL:
human STS SHGC-11774.

Score = 1643, P = 8.0e-67, identities = 345/355

Entry HS513217 from database EMBL:
human STS SHGC-14656.

Score = 1193, P = 5.8e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 263 bp to 1885 bp; peptide length: 541
Category: similarity to known protein

```

1  MLSHTVRKHL RKTRLELLHK EYEDEIDCLQ KEVEELKSKN LSLESQIKAI
51  LDPLTLVQGN QNEDKHLVTD NPSIINPETV AEWKKKLRTA NEIYEKVKDD
101 VDKLKEANKK LKLENGGLVR ENLRLKAEVD NRSPOKEGRF AVAALQSKVE
151 QYERETNRLK KALERSDKYI EELESQVAQL KNSSEEKEAM NSICQTALSA
201 DGKGSKGSEE DVVSKNQGDS ARKQPGSSTS SSSHLAKPSS SRLCOTSSAR
251 QESTSKADLN CSKNKDLYQE QVEVMDLVD TSMOTYLERE WGNKPSDCVP
301 YKDEELYDFP APCTPLSLSC LQLSTPENRE SSVVQAGGSK KHSNHLRLV
351 FDDFCDSSNV SNKDSSEDDI SRSENEKKSE CFSSTKTGEW DCCSTSYAQN
401 LDFESSEGNT IANSVGEISS KLSEKSGLC SKRLNSIRSF EMNRTRTSSE
451 ASMDAAYLDK ISELDMMSE SDNSKSPCNN GFKSLDLGL SKSSQGSEFL
501 EEPDKLEEKT ELNLSKGLT NDQLENGSEW KPTSFPSPLS I

```

BLASTP hits

Entry A42771 from database PIR:
reticulocyte-binding protein 1 - Plasmodium vivax
Score = 127, P = 3.7e-08, identities = 68/300, positives = 145/300

Entry RBP1 PLAVB from database SWISSPROT:
RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
Score = 127, P = 3.9e-08, identities = 68/300, positives = 145/300

Entry MMDSPPG_1 from database TREMBL:
gene: "DSPP"; product: "dentin sialophosphoprotein"; Mus musculus DSPP
gene
Score = 160, P = 5.2e-08, identities = 87/373, positives = 146/373

Alert BLASTP hits for DKFZphfbr2_62n10, frame 2

No Alert BLASTP hits found

PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00001	443->447	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	526->530	ASN_GLYCOSYLATION	PDOC00001
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	156->159	PKC_PHOSPHO_SITE	PDOC00005
PS00005	166->169	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	361->364	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	419->422	PKC_PHOSPHO_SITE	PDOC00005
PS00005	423->426	PKC_PHOSPHO_SITE	PDOC00005
PS00005	431->434	PKC_PHOSPHO_SITE	PDOC00005
PS00005	436->439	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	324->328	CK2_PHOSPHO_SITE	PDOC00006
PS00006	361->365	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	373->377	CK2_PHOSPHO_SITE	PDOC00006
PS00006	414->418	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	462->466	CK2_PHOSPHO_SITE	PDOC00006
PS00006	469->473	CK2_PHOSPHO_SITE	PDOC00006
PS00007	294->302	TYR_PHOSPHO_SITE	PDOC00007
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	226->232	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	408->414	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	517->523	MYRISTYL	PDOC00008
PS00013	310->321	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_62n10.2)

DKFZphfbr2_62o17

group: metabolism

DKFZphfbr2_62o17.2 encodes a novel 282 amino acid protein with weak similarity to the apolipoprotein E receptor.

The new protein contains a leucine zipper for protein-protein interaction, and three LDL-receptor class A domain (LDLRA_1) patterns. In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands.

The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins

similarity to apolipoprotein E receptor

complete cDNA, complete cds, start at Bp 56 matches kozak consensus
ANCatg EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1260 bp

Poly A stretch at pos. 1240, polyadenylation signal at pos. 1218

```
1 GGGGGATAAG AGAGCGGTCT GGACAGCGCG TGGCCGGCGC CGCTGTGGGG
51 ACAGCATGAG CGGCGGTGG ATGGCGCAGG TTGGAGCGTG GCGAACAGGG
101 GCTCTGGGCC TGGCGCTGCT GCTGCTGCTC GGCCCTCGGAC TAGGCCTGGA
151 GGCCGCGCGC AGCCCGCTTT CCACCCGAC CTCTGCCAG GCCCGAGGCC
201 CCAGCTCAGG CTCTGCCCCA CCCACCAAGT TCCAGTGCCG CACCAGTGGC
251 TTATGCGTGC CCCTCACCTG GCGCTGCGAC AGGGACTTGG ACTGCAGCGA
301 TGGCAGCGAT GAGGAGGAGT GCAGGATTGA GCCATGTACC CAGAAAGGGC
351 AATGCCCAAC GCCCCCTGGC CTCCCCTGCC CCTGCACCGG CGTCAGTGAC
401 TGCTCTGGGG GAACTGACAA GAAACTGCGC AACTGCAGCC GCCTGGCCTG
451 CCTAGCAGGC GAGTCCGTT GCACGCTGAG CGATGACTGC ATTCCACTCA
501 CGTGGCGCTG CGACGGCCAC CCAGACTGTC CCGACTCCAG CGACGAGCTC
551 GGCTGTGGAA CCAATGAGAT CCTCCCGGAA GGGGATGCCA CAACCATGGG
601 GCCCCCTGTG ACCCTGGAGA GCGTCACCTC TCTCAGGAAT GCCACAACCA
651 TGGGGCCCCC TGTGACCCTG GAGAGTGTCC CCTCTGTCGG GAATGCCACA
701 TCCTCCTCTG CCGGAGACCA GTCTGGAAGC CCAACTGCCT ATGGGGTTAT
751 TGACAGTGCT GCGGTGCTCA GTGCAAGCCT GGTCAACCGC ACCCTCCTCC
801 TTTTGTCTGT GCTCCGAGCC CAGGAGCGCC TCCGCCCACT GGGGTTACTG
851 GTGGCCATGA AGGAGTCCCT GCTGCTGTCA GAACAGAAGA CCTCGCTGCC
901 CTGAGGACAA GCACTTGCCA CCACCGTCAC TCAGCCCTGG GCGTAGCCGG
951 ACAGGAGGAG AGCAGTGATG CGGATGGGTA CCGGGGCACA CCAGCCCTCA
1001 GAGACCTGAG CTCCTCTGGC CACGTGGAAC CTCGAACCCG AGCTCCTGCA
1051 GAACTGGCCC TGGAGATTGA GGGTCCCTGG ACACTCCCTA TGGAGATCCG
1101 GGGAGCTAGG ATGGGGAACC TGCCACAGCC AGAACCGAGG GGCTGGCCCC
1151 AGGCAGCTCC CAGGGGGTAG GACGGCCCTG TGCTTAAGAC ACTCCTGCTG
1201 CCCCCTCTGA GGGTGGCGAT TAAAGTTGCT TCACATCCTC AAAAAAAAAA
1251 AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 56 bp to 901 bp; peptide length: 282
Category: similarity to known protein
Classification: unset
Prosites motifs: LDLRA_1 (67-90)
LDLRA_1 (67-90)
LDLRA_1 (145-168)

LEUCINE_ZIPPER (17-39)

```

1 MSGGWMAQVG AWRTGALGLA LLLLGLGLG LEAAASPLST PTSQAAGPS
51 SGSCPTKFKQ CRTSGLCVPL TWRCRDRLDC SDGSDEEECR IEPCTQKGQC
101 PPPPGLPCPC TGVSDCSGGT DKRLNCSRL ACLAGELRCT LSDDCIPLTW
151 RCDGHPDCPD SSDELGCCTN EILPEGDATT MGPPVTLESV TSLRNATTMG
201 PPVTLESVPS VGNATSSSAG DQSGSPTAYG VIAAAVLSA SLVTATLLLL
251 SWLRAQERLR PLGLLVAMKE SLLSEQKTS LP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62o17, frame 2

TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene., N = 1, Score = 733, P = 1.5e-72

PIR:JE0237 apolipoprotein E receptor 2 precursor - mouse, N = 2, Score = 290, P = 1.1e-26

TREMBL:HS275190.1 product: "apolipoprotein E receptor 2 906"; H.sapiens mRNA for apolipoprotein E receptor 2, N = 1, Score = 279, P = 1.8e-23

>TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene.
Length = 260

HSPs:

Score = 733 (110.0 bits), Expect = 1.5e-72, P = 1.5e-72
Identities = 157/276 (56%), Positives = 178/276 (64%)

```

Query:      6 MAQVGAWRTGALGLALLLLGLGLGLEAAASPLSTPTSQAAGPSSGSCPTKFCRTSG 65
             MA+ GA R ALGL L LL GL GLEAA +P T Q +G + SCP FQC TSG
Sbjct:      1 MARGGAGRAVALGLVLRLLFGLRTGLEAAPAPAHNT--RVQVSGSRADSCPTDTFQC LTSG 58

Query:      66 LCVPLTWRCRDRLDCSDGSDEEECRIEPCTQKGQC P P P P P L P C P C T G V S D C S G G T D K K L R 125
             CVPL+WRCD D DCSDGSDEE+CRIE C Q GQC P LPC C +S CS +DK L
Sbjct:      59 YCVPLSWRCGDQDCSDGSDEEDCRIESCANGQCQPQSALPCSDNISGCSDVSDKNL- 117

Query:      126 NCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGCCTNEILPEGDATTMGPPV 185
             NCSR C EL C L D CIP TWRCDGHPDC DSSDEL C T+
Sbjct:      118 NCSRPPCQESLHCILDDVCIPHTWRCDGHPDCLDSSDELSCDTD-----T 163

Query:      186 TLESVTSRLRNATTMGPPVTLESVPSVGNATSSSAGDQSGSPTAYGVIAAAVLSASLVTA 245
             ++ + NATT T+E+ S N T +SAGD S +P+AYGVIAAA VLSA LV+A
Sbjct:      164 EIDKIFQEENATTTRISTTMENETSFRNVTFTSAGDSSRNPSAYGVIAAGVLSAILVSA 223

Query:      246 TLLLSWLRAQERLRPLGLLVAMKESLLSEQKTS L 281
             TLL+L LR Q L P GLLVA+KESLLSE+KTS L
Sbjct:      224 TLLILLRLRGQYLPPPGLLVAVKESLLSERKTS L 259

```

Pedant information for DKFZphfbr2_62o17, frame 2

Report for DKFZphfbr2_62o17.2

```

[LENGTH]      282
[MW]           28991.19
[pI]           4.61
[HOMOL]        TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility
                complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein,
                BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes,
                complete cds; Sacm21 gene, partial cds; and unknown gene. 5e-55
[BLOCKS]       BL01209 LDL-receptor class A (LDLRA) domain proteins
[SCOP]         diaj__ 7.11.1.1.1 Ligand-binding domain of low-density lipoprotein 2e-10

```


[illegible]

```

PS01209      67->90      LDLRA_1      PDOC00929
PS01209      67->90      LDLRA_1      PDOC00929
PS01209      145->168    LDLRA_1      PDOC00929
PS00029      17->39      LEUCINE ZIPPER PDOC00029

```

HMM_NAME	TNFR/NGFR cysteine-rich region	
HMM	*CpeGtYtD.WNHvpqClpC.trCePEMGQYmvpPCTwTQNT.VC*	
	CP+ ++ + + C+P RC+ ++ +C + ++ +C	
Query	54 CPPTKFCQRTS--GLCVPLTWRCDR--DL----DCSDGSDEEEEC	89

HMM_NAME Low-density lipoprotein receptor domain class A

HMM *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPENChp*
C P +FQC+++ C+P+ W+CD D DC D+SDE E+C+

Query 52 GSCP-PTKFQCRTSG-LCVPLTWRCDRDLDCSDGSDE--EECRI 91

54.99 (bits) f: 130 t: 169 Target: dkfzphfbr2_62o17.2 similarity to apolipoprotein E receptor

Alignment to HMM consensus:

Query *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPENChp*
C + E +C + CIP+ W+CDG PDC D SDE ++C+

dkfzphfbr2 130 LACL-AGELRCTLSD-DCIPLTWRCGHPCPDSSDE--LGCGT 169

DKFZphfbr2_64a15

group: nucleic acid management

DKFZphfbr2_64a15 encodes a novel 255 amino acid protein with strong similarity to inorganic pyrophosphatases

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity.

The new protein can find application as a new enzyme for biotechnologic processes.

strong similarity to inorganic pyrophosphatases

unspliced Intron 212-256 see EST HS1190948

Sequenced by Qiagen

Locus: unknown

Insert length: 1188 bp

Poly A stretch at pos. 1170, polyadenylation signal at pos. 1151

```

1 GGGGGTTGGG GACCAGTGCA GGGACCGGGT CGCGCCGTGC TATGGCCCTG
51 TACCACACTG AGGAGCGCGG CCAGCCCTGC TCGCAGAATT ACCGCCTCTT
101 CTTTAAGAAT GTAACCTGGT ACTACATTTC CCCCTTTTCAT GATATTCCTC
151 TGAAGGTGAA CTCTAAAGAG GACACTGAGG CTCAAGGCAT TTTTATAGAC
201 TTGTCTAAGA TCTGGAAAAT GGCATTCCCTA TGAAGAAAGC ACGAAATGAT
251 GAATATGAGA ATCTGTTTAA TATGATTGTA GAAATACCTC GGTGGACAAA
301 GGCTAAAATG GAGATTGCCA CCAAGGAGCC AATGAATCCC ATTAACAAT
351 ATGTAAAGGA TGGAAAGCTA CGCTATGTGG CGAATATCTT CCCTTACAAG
401 GGTATATAT GGAATTATGG TACCCTCCCT CAGACTTGGG AAGATCCCCA
451 TGA AAAAGAT AAGAGCACGA ACTGCTTTGG AGATAATGAT CCTATTGATG
501 TTTGCGAAAT AGGCTCAAAG ATTCTTTCTT GTGGAGAAGT TATTTCATGTG
551 AAGATCCTTG GAATTTTGGC TCTTATTGAT GAAGGTGAAA CAGATTGGAA
601 ATTAATTGCT ATCAATGCGA ATGATCCTGA AGCCTCAAAG TTTTCATGATA
651 TTGATGATGT TAAGAAGTTC AAACCGGGTT ACCTGGAAGC TACTCTTAAT
701 TGGTTTAGAT TATGTAAGGT ACCAGATGGA AAACCAGAAA ACCAGTTTGC
751 TTTTAATGGA GAATTCAAAA ACAAGGCTTT TGCTCTTGAA GTTATTAAAT
801 CCACTCATCA ATGTTGGAAG GCATTGCTTA TGAAGAAGTG TAATGGAGGA
851 GCTACAAATT GCACAAACGT GCAGATATCT GATAGCCCTT TCCGTGTCAC
901 TCAAGAGGAA GCAAGATCAT TAGTTGAATC GGTATCATCT TCACCAATA
951 AAGAAAGTAA TGAAGAAGAG CAAGTGTGGC ACTTCCTTGG CAAGTGATTG
1001 AAACATCTGA AATTCTGCTG TCAAGATTCC CATCTCTAAG GACTCCAAGA
1051 CTCTTTTCC CCAAGTGCTA GAGACAAGGG GGTCTATGAG CATTTACTGA
1101 CTTCTGTGTA AAACCTTCATT TTTTCAAAC TTTTGAGCTA TGCAATATAT
1151 AAATAAACAG TAAGAATTTT AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HSPPASEMR from database EMBL:
H.sapiens partial mRNA for pyrophosphatase.
Score = 1706, P = 1.6e-70, identities = 342/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 230 bp to 994 bp; peptide length: 255
Category: strong similarity to known protein
Classification: unset
Prosites motifs: PPASE (85-92)

1 MKKARND EYE NLFNMIVEIP RWTAKMEIA TKEPMNPIQ YVKDGKLRVY
 51 ANIFPYKGYI WNYGTLPTW EDPHEKDKST NCFGDNNDPID VCEIGSKILS
 101 CGEVIHV KIL GILALIDEGE TDWKLIAINA NDPEASKFHD IDDVKKFKPG
 151 YLEATLNWFR LCKVPDGKPE NQFAFNGEFK NKAFAL EVIK STHQCWKALL
 201 MKNCNGGATN CTNVQISDSP FRCTQEEARS LVESVSSSPN RESNEEEQVW
 251 HFLGK

BLASTP hits

Entry IPYR KLULA from database SWISSPROT:
 INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-
 HYDROLASE) (PPASE).
 Score = 689, P = 6.0e-68, identities = 128/248, positives = 170/248

Entry A45153 from database PIR:
 inorganic pyrophosphatase (EC 3.6.1.1) - bovine
 Score = 862, P = 2.8e-86, identities = 146/226, positives = 190/226

Entry AF085600.1 from database TREMBLNEW:
 gene: "Nurf-38"; product: "inorganic pyrophosphatase NURF-38";
 Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38)
 gene, complete cds.
 Score = 731, P = 2.1e-72, identities = 134/248, positives = 177/248

Entry PWB1 from database PIR:
 inorganic pyrophosphatase (EC 3.6.1.1) - yeast (Saccharomyces
 cerevisiae)
 Score = 688, P = 7.7e-68, identities = 133/251, positives = 174/251

Alert BLASTP hits for DKFZphfbr2_64a15, frame 2

SWISSPROT:IPYR DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
 (PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 731, P =
 2.4e-72

>SWISSPROT:IPYR DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
 PHOSPHO- HYDROLASE) (PPASE).
 Length = 290

HSPs:

Score = 731 (109.7 bits), Expect = 2.4e-72, P = 2.4e-72
 Identities = 134/248 (54%), Positives = 177/248 (71%)

Query: 7 DEYENLFNMIVEIPRWTAKMEIATKEPMNPIQYVKDGKLRVYANIFPYKGYIWNNGTL 66
 +E + ++NM+VE+PRWT AKMEI+ K PMNPIQ +K GKL+VAN FP+KGYIWNNG L
 Sbjct: 40 NEEKTIYNMVVEVPRWTNAKMEISLKTPMNPIQDIKKGLRFVANCFFHKGYIWNNGAL 99

Query: 67 PQTWEDPHEKDKSTNCFGDNNDPIDVCEIGSKILSCGEVIHV KILGILALIDEGETDWKLI 126
 PQTWE+P + ST C GDNDPIDV EIG ++ G+V+ VK+LG ALIDEGETDWK+I
 Sbjct: 100 PQTWENPDHIEPSTGCKGDNNDPIDVIEIGYRVAKRGDVLKVKVLGQFALIDEGETDWKII 159

Query: 127 AINANDPEASKFHDIDDVKKFKPGYLEATLNWFR LCKVPDGKPE NQFAFNGEFKNKAFAL 186
 AI+ NDP ASK +DI DV ++ PG L AT+ WF++ K+PDGKPE NQFAFNG+ KN FA
 Sbjct: 160 AIDVNDPLASKVNDIADVQYFPGLLRATVEWFKIYKIPDGKPE NQFAFNGDAKNADFAN 219

Query: 187 EVIKSTHQCWKALLMKNCNGGATNCTNVQISDSPFRCTQEEARS-LVESVSSSPNKESNE 245
 +I TH+ W+ L+ ++ G+ + TN+ +S +EEA L E+ +E ++
 Sbjct: 220 TIIAETHKFWQNLVHQSPASGSISTTNITNRNSEHVIPKEEA EKILAEAPDGGQVEEVSD 279

Query: 246 EEQVWHFL 253
 WHF+
 Sbjct: 280 TVDTWHFI 287

Peptide information for frame 3

ORF from 42 bp to 230 bp; peptide length: 63
 Category: strong similarity to known protein
 Classification: unset

1 MALYHTEERG QPCSQNYRLF FKNVTGHIYS PFHDIPLKVN SKEDTEAQGI
 51 FIDLSKIWKM AFL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64a15, frame 3

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
 (PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 118, P =
 8.8e-07

PIR:A45153 inorganic pyrophosphatase (EC 3.6.1.1) - bovine, N = 1,
 Score = 113, P = 3.1e-06

TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase";
 Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds., N
 = 1, Score = 106, P = 1.8e-05

>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
 PHOSPHO- HYDROLASE) (PPASE).
 Length = 290

HSPs:

Score = 118 (17.7 bits), Expect = 8.8e-07, P = 8.8e-07
 Identities = 23/43 (53%), Positives = 29/43 (67%)

Query: 1 MALYHTEERGQPCSONYRLFFKNVTGHIYISPFHDIPLKVNSE 43
 MALY T E+G S +Y L+FKN G+ ISP HDIPL N ++
 Sbjct: 1 MALYETVEKGAKNSPSYSLYFKNKCGNVISPMHDIPLYANEK 43

Pedant information for DKFZphfbr2_64a15, frame 2

Report for DKFZphfbr2_64a15.2

[LENGTH] 255
 [MW] 29177.34
 [pI] 5.67
 [HOMOL] TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase"; Homo
 sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds. 2e-93
 [FUNCAT] 01.04.01 phosphate utilization [S. cerevisiae, YBR011c] 9e-73
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR011c] 9e-73
 [FUNCAT] 02.99 other energy generation activities [S. cerevisiae, YMR267w] 1e-58
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR267w] 1e-58
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 genitalium, MG351] 1e-06
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0124] 2e-06
 [BLOCKS] BL00387D
 [BLOCKS] BL00387C
 [BLOCKS] BL00387B
 [BLOCKS] BL00387A
 [SCOP] dlwgja_2.29.5.1.1 Inorganic pyrophosphatase [baker's yeas 1e-113
 [EC] 3.6.1.1 Inorganic pyrophosphatase 7e-92
 [PIRKW] mitochondrion 3e-57
 [PIRKW] hydrolase 7e-92
 [PIRKW] homodimer 2e-71
 [SUPFAM] inorganic pyrophosphatase 7e-92
 [PROSITE] PPASE 1
 [KW] Alpha_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 6.27 %

SEQ MKKARNDEYENLFNMIVEIPRWTAKMEIATKEPMNPIKQYVRDGLRYVANIFPYKGYI
 SEG
 lhukBEGGGCEEEEEETTTbCBCEBETTTTTTCEEECEETTEECBCCBBTBTtBT

SEQ WNYGTLPQTWEDPHEKDKSTNCFGNDPIDVCEIGSKILSCGEVIHVILGILALIDEGE
 SEG
 lhukB CEEEEETTTTCBTTTTEETTTTECCCBCEEECCCCCTTTTEEEEEEEEEETTTT

SEQ TDWKLIAINANDPEASKFHDIDDVKKFKPGYLEATLNWFRICKVDPGKPNQFAFNGEFK
 SEG
 lhukB CEEEEEEETTTTGGGCCCHHHHHHTTTTHHHHHHHHHHCGGGCCCCBCGGGCCB

SEQ NKAFALEVIKSTHCWKALLMKNCGGATNCTNVQISDSPFRCTQEEARSLVESVSSSPN
 SEGxxxxxxx
 lhukB CHHHHHHHHHHHHHHHHHHCTTTTTTCCCBTTTTTTT

DKFZphfbr2_64c16

group: brain derived

DKFZphfbr2_64a16.2 encodes a novel 101 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: /map="745_A_2; 756_F_2; 842_C_2"

Insert length: 1866 bp

Poly A stretch at pos. 1848, polyadenylation signal at pos. 1829

```
1 GGGCGCGGCG CCGGAGGAGG AAGTGGTGAG GTTGTGCTC CTTACGCGC
51 TATCGCTGGC TCTTGGGGCG CAGAGAGGGG CCGCAGTCTC CGCGGCTGCG
101 TCGAGCTCCC TTGCAGTCCC CTCCATGTTT CCGGCGGCCA CTACTCCCCT
151 TCCTAAGGCC GCCGCTTACC CCGGGGTCTA TGGAGTAAT GGAAGGACCC
201 CTCACCTGG CTCATCAACA GAGCAGACGA GCAGACCGTT TATTAGCTGC
251 AGGCCAATAC GAAGAGGCTA TTTCTTGTC CAAAAAGGCT GCAGCATATC
301 TTTCTGAAGC CATGAAGCTG ACACAGTCAG AGCAGGCTCA TCTTCACTG
351 GAATTGCAAA GGGATAGCCA TATGAAACAG CTCCTCCTCA TCCAAGAGAG
401 ATGGAAGAGG GCCCAGCGTG AAGAAAGATT GAAAGCCCAG CAGAACACAG
451 ACAAGGATGC AGCTGCCCAT CTTAGACAT CTCACAAACC CTCTGCAGAG
501 GATGCAGAGG GCCAGAGTCC CTTTCTCAG AAGTACAGCC CTTCCACAGA
551 GAAATGCCTG CCTGAGATTC AGGGGATCTT TGACAGGGAT CCAGACACAC
601 TACTTTATTT ACTTCAGCAA AAGAGTGAGC CAGCAGAGCC ATGTATTGGA
651 AGCAAGAGCC CAAAAGATGA TAAACAATT ATAGAGGAGC AGGCAACCAA
701 AATTGCAGAT TTGAAGAGGC ATGTGGAATT CTTGTGGCT GAGAATGAAA
751 GATTAAGGAA AGAAAAATAA CAACTAAAGG CTGAAAAGGC CAGACTTCTA
801 AAAGGTCCAA TAGAAAAGGA GCTGGATGTA GATGCTGATT TTGTAGAAAC
851 GTCAGAGTTA TGGAGCTTGC CACCACATGC AGAACTGCT ACAGCCTCCT
901 CAACCTGGCA GAAGTTCGCA GCAAACTAGT GGAAGCCAA GGACATTCCA
951 ATCCCCAATC TTCTCCTCTT GGAATTTCCA TCTCCAGAAC TTCTCTTAT
1001 GGAGCTCTCT GAGGATATTC TGAAAGGACT TATGAATAAT TAAATGGAA
1051 GGCCACAGAA AAGGGGAAAA GAGGAATAA TACAGTAATC GTTAATCCAG
1101 CAAAAGAGAA TGAAAAGGGA AAACCACATA GAAGGGTAAT CCCGGAAATG
1151 CTTCACTCTG TGGACTGTGG GAGCAGAGGC ATTGCCAGGA CTTGGGAAAC
1201 AGTCACTGTG AATGCGCTG CGTATCTCAT TCACTCACTT CAGCTAATGA
1251 CTCCGACTTG GCAGACGCTA AACTCATGGA GGTTCGGTTT CTCCTGATAC
1301 AAACCAATG GCTACCTGGA AGAATTTCTT TCAAGCAACA GTTATTTTTC
1351 TTATCTTCAG GGTAAATATG TATAAAGTT ATGTGTAATT AATCTATAAT
1401 GCCATAATG ATAATGCAAA ACCTAAATAA TATGGTGGCC GGAGGGGCTG
1451 CCTTATATTT GAAACATGCT TTCTATCATG CATTGACTGT ATGCATTTTG
1501 TTAATGCACA TTCTGTTTGT TTAAGGTGTG TGAGATACAC ACCTTTCTAG
1551 ATGAAACTAT ATGTGCCACA CTTTGCACTA CTCATAATGA TAACCTCAAG
1601 ACTATCAGAA GAAATATTTA AATTTCCATT TTATGAAGAA AGGAACCAAA
1651 TTATTATGCT TTTTAAACA AATTACCAGT TTACATAATT AATCAGGGTG
1701 CATTTTAAGT TCTAACTTCG TTTATTGTAT AATGCATCAT TTGAAAATAC
1751 CAAGGAGGAA ATACCCTTTG TTTTAAATGA TGCAAGAGTG GACGTAATGC
1801 TAGTTGGCAG TATTTTATTG TAAGAAATCA ATAAAGTAAT TGTGTTTTAA
1851 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS286143 from database EMBL:

human STS WI-6844.

Score = 1460, P = 3.4e-61, identities = 292/292

Medline entries

No Medline entry

ORF from the beginning to 304 bp; peptide length: 102
Category: questionable ORF
Classification: unset

1 GAAPEEEVVR LLLLQRLSLA LGAQRGA AVS AAASSSLAVP SMFPGATTPL
51 PKAAAYPGVY GSGRTPPQG SSTEQTSRPF ISCRQIRRGY FLSQKGC SIS
101 F

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64c16, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 180 bp to 1040 bp; peptide length: 287
Category: putative protein
Classification: unset
Prosites motifs: LEUCINE ZIPPER (178-200)
LEUCINE ZIPPER (185-207)

```

1 MEVMEGPLNL AHQQSRRADR LLAAGKYEEA ISCHKKAAAY LSEAMKLTQS
5 EQAHLSELEQ RDSHMKQLLL IQRWKRAQPT EERGLKAPQNT LKDAAAHLQT
101 SHKPSAEDAQ GQSPLSQKYS PQTECKLPEI QGFDRDPDT DDLYLLQQSE
151 PAEPCIGSKA PKDDKTIIEE QATKIADLKR HVEFLVAENE RLRENKQKQ
201 AEKARLLKGP ITELDDVDAD FVTSELSWL PPHAETATAS STWQKFAANT
251 GKAKDIPN LPPLDFPSPE LPMLESEDI LKGLMNN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64c16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64c16, frame 2

Report for DKF2phfbr2_64c16.2

```
[LENGTH]      101
[MW]           10469.94
[pI]           10.18
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      29.70 %
```

```
SEQ      GAAPEEEEVRLLLLQRLSLALGAQRGAAVSAAASSSLAVPSMFPGATTPLPKAAAYPGVY
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccc
```

```

SEQ      GSNGRTPQPGSSTEQTSRPFISCRQIRRGYFLSQKGCSISF
SEG      .....
PRD      cccccccccccccccccccccccccchhhhhcccccccccccccc

```

(No Prosite data available for DKFZphfbr2_64c16.2)

(No Pfam data available for DKFZphfbr2_64c16.2)

Pedant information for DKF2phfbr2_64c16, frame 3

Report for DKFZphfbr2_64c16.3

```

{LENGTH}      287
{MW}           32343.79
{pI}           5.61
{PROSITE}      LEUCINE_ZIPPER 2
{KW}           All_Alpha
{KW}           COILED_COIL      14.98 %

```

```

SEQ  MEVMEGFLNLAHQQSRRADRLLAAGKYEEAISCHKKAAAYLSEAMKLTQSEQAHLSELEQ
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

```

```

SEQ  RDSHMKQLLLIQERWKRAQREERLKAQNTDKDAAHLQTSKKPSAEDAEGQSPLSQKYS
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

```

```

SEQ  PSTEKCLPEIQIFDRDPDTLLYLLQKSEPAEPCIGSKAPKDDKTIIEEQATKIADLKR
PRD  cccccccchhhhhcccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCC

```

```

SEQ  HVEFLVAENERLRKENKQLKAEKARLLKGPIEKELDVDADFVETSELWSLPPHAETATAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

```

```

SEQ  STWQKFAANTGKAKDIPINLPPLDFPSPPELPLMELSEDILKGLMNN
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

```

Prosites for DKFZphfbr2_64c16.3

```

PS00029  178->200  LEUCINE_ZIPPER      PDOC00029
PS00029  185->207  LEUCINE_ZIPPER      PDOC00029

```

(No Pfam data available for DKFZphfbr2_64c16.3)

DKFZphfbr2_64c4

group: brain derived

DKFZphfbr2_64c4 encodes a novel 467 amino acid protein with similarity to A. thaliana T08I13.5

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A. thaliana T08I13.5

complete cDNA, complete cds, EST hits
on genomic level encoded by AC005043 11 exons

Sequenced by Qiagen

Locus: unknown

Insert length: 1559 bp

Poly A stretch at pos. 1540, no polyadenylation signal found

```

1 TGGGACCGCC GGAAGTTTCT GCCGCGGCTT TCGGGGACG GGGGAGTGGT
51 AGTGGGGGCT GCAGCTGCCG GACCCAGGCG CGATGGCTAC GGGCGCGGAT
101 GTACGGGACA TTCTAGAACT CGGGGGTCCA GAAGGGGATG CAGCCTCTGG
151 GACCATCAGC AAGAAGGACA TTATCAACCC GGACAAGAAA AAATCCAAGA
201 AGTCCTCTGA GACACTGACT TTCAAGAGGC CCGAGGGCAT GCACCGGGAA
251 GTCTATGCCT TGCTCTACTC TGACAAGAAG GATGCACCCC CACTGCTACC
301 CAGTGACACT GGCCAGGGAT ACCGTACAGT GAAGGCCAAG TTGGGCTCCA
351 AGAAGGTGCG GCCTTGGAAG TGGATGCCAT TCACCAACCC GGGCCGCAAG
401 GACGGAGCAA TGTTCTTCCA CTGGCGACGT GCAGCGGAGG AGGGCAAGGA
451 CTACCCCTTT GCCAGGTTCA ATAAGACTGT GCAGGAGCCT GTGTACTCGG
501 AGCAGGAGTA CCAGCTTTAT CTCCACGATA ATGCTTGGAC TAAGGCAGAA
551 ACTGACCACC TCTTTGACCT CAGCCGCCGC TTTGACCTGC GTTTTGTGTG
601 TATCCATGAC CGGTATGACC ACCAGCAGTT CAAGAAGCGT TCTGTGGAAG
651 ACCTGAAGGA GCGGTACTAC CACATCTGTG CTAAGCTTGC CAACGTGCGG
701 GCTGTGCCAG GCACAGACCT TAAGATACCA GTATTGATG CTGGGCACGA
751 ACGACGGCGG AAGGAACAGC TTGAGCGTCT CTACAACCGG ACCCCAGAGC
801 AGGTGGCAGA GGAGGAGTAC CTGCTACAGG AGCTGCGCAA GATTGAGGCC
851 CGGAAGAAGG AGCGGGAGAA ACGCAGCCAG GACCTGCAGA AGCTGATCAC
901 AGCGGCAGAC ACCACTGCAG AGCAGCGGCG CACGGAACGC AAGGCCCCCA
951 AAAAGAAAGT ACCCCAGAAA AAGGAGGCTG AGAAGCCGCG TGTTCTTGAG
1001 ACTGCAGGCA TCAAGTTTCC AGACTTCAAG TCTGCAGGTG TCACGCTGCG
1051 GAGCCAAACG ATGAAGCTGC CAAGCTCTGT GGGACAGAAG AAGATCAAGG
1101 CCCTGGAACA GATGCTGCTG GAGCTTGGTG TGGAGCTGAG CCCGACACCT
1151 ACGGAGGAGC TGGTGACAT GTTCAATGAG CTGCGAAGCG ACCTGGTGCT
1201 GCTCTACGAG CTCAAGCAGG CCTGTGCCAA CTGCGAGTAT GAGCTGCAGA
1251 TGCTGCGGCA CCGTCATGAG GCACTGGCCC GGGCTGGTGT GCTAGGGGGC
1301 CCTGCCACAC CAGCATCAGG CCCAGGCCCG GCCTCTGCTG AGCCGGCAGT
1351 GTCTGAACCC GGAATTGGTC CTGACCCCAA GGACACCATC ATTGATGTGG
1401 TGGGCGCACC CCTCAGGCC AATTCGAGAA AGCGACGGGA GTCGGCCTCC
1451 AGCTCATCTT CCGTGAAGAA AGCCAAGAAG CCGTGAGAGG CCCCACGGGG
1501 TGTGGGCGAC GCTGTTATGT AAATAGAGCT GCTGAGTTGG AAAAAAAAAA
1551 AAAAAAAAAA

```

BLAST Results

Entry AC005043 from database EMBL:
Homo sapiens clone NH0576N21; HTGS phase 1, 5 unordered pieces.
Score = 1506, P = 4.6e-244, identities = 316/330

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 83 bp to 1483 bp; peptide length: 467


```

1 MATGADVDRDI LELGGPEGDA ASGTISKKDI INPDKKSKK SSETLTFKR
51 EGMHREVVAL LYSDKKDAPL LPSDPTGGY RTVKAPKLGSK KVRPWKWMF
101 TNPARKDGDAM FFLHWRRAEE KVDYDFARFN QTFQKPVSE
151 AWTKAETHDL FDSRRFDLR FKVIHBDYDH KQVQKSVED
201 KLANVRAPVG TOLKI PVFDA GHERRRKEQL ERLYNRTPE
251 LRKIEARKKE REKRSQDLQK LITAOTTA E QRRTERKAP
301 KPAPVETAGI KFPDFKSAGV TIRSQRMKLK SSVGQKKIKA
351 ELSPTTEEL VHMFNELRSD LVLLYELKQA CANCEYELQM
401 AGVLGGPPAT ASGPGPASAE PAVSEPLGP DPKDTIIDV
451 RRESAGSSSS VKKAKVP

```

Entry ATAC2337 5 from database TREMBLNEW:
gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC T08I13
genomic sequence, complete sequence.
Score = 340, P = 2.6e-30, identities = 115/374, positives = 176/374

Entry S64291 from database PIR:
hypothetical protein YGR002c - yeast (*Saccharomyces cerevisiae*)
Score = 202, P = 2.8e-13, identities = 71/260, positives = 124/260

No Alert BLASTP hits found

Report for DKFZphfbr2 64c4.2

```
[LENGTH]          467
[MW]               53007.60
[PI]               9.51
[HOMOL]            TREMBL:ATAC2337_5 gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC
T08I13 genomic     sequence, complete sequence. 4e-29
[FUNCAT]           99 unclassified proteins          [S. cerevisiae, YGR002c] 1e-19
[PROSITE]          MYRISTYL 1
[PROSITE]          CAMP_PHOSPHO_SITE 4
[PROSITE]          CK2_PHOSPHO_SITE 10
[PROSITE]          TYR_PHOSPHO_SITE 3
[PROSITE]          GLYCOSAMINOGLYCAN 1
[PROSITE]          PKC_PHOSPHO_SITE 12
[PROSITE]          ASN_GLYCOSYLATION 1
[KW]               All Alpha
[KW]               LOW COMPLEXITY 20.13 %
```

```
SEQ      MATGADVDRDILELGGPEGDAASGTISKDDIINPDKKSKKSSSETLTFKRPEGMHREYVAL
SEG      .....xxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccceeeeeeeecccccccccccccccccccccccccccccccccchhhhhhhh

SEQ      LYSDDKADAPLLPSDTGGQGYRTVAKLGSKKVRPWKMFPFTNPARKDGAMFFHWRRAAE
SEG      .....
PRD      hhhccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhc

SEQ      GKDYPFARFNKTVOEPVYSEQEYQLYLHDNAWTKAETHLFDLSRRFDFLRVVIIHdrydh
SEG      .....
PRD      ccccccccccccccccchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhccceeeeecccc

SEQ      QQFKKSrVEDLKERYHYICAKLANVRAPGTDLKIPVFDAGHERRKEQLERLYNRTPEQ
SEG      .....
PRD      chhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhc

SEQ      VAEeEYLLQELRKIEARKKEREKRSQDLQKLITAADTTAEQRRTERKAPKKKLPOKKEAE
SEG      .....xxxxxxxxxxxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhh

SEQ      KPAAVPETAGIKFPDFKSAGVTILSRQMRLPSSVGQKKIKALEQMLLELGVELSPPTTEEL
SEG      xxx.....
```


Prosites for DKF2phfbr2_64c4.2

(No Pfam data available for DKFZphfbr2_64c4.2)

DKFZphfbr2_64h6

group: brain derived

DKFZphfbr2_64h6 encodes a novel 176 amino acid protein with similarity to predicted yeast proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to S.pombe SPBC337.09 and S.cerevisiae YER044c

complete cDNA, complete cds accoring to YER044c/SPBC337.09,
start at Bp 111, EST hits

Sequenced by Qiagen

Locus: /map="14"

Insert length: 1212 bp

Poly A stretch at pos. 1192, polyadenylation signal at pos. 1168

```

1 GGGCTGGAGC TGCTCTGGGG GAGCTTGTTT GCGGCAGCGG CTGCTGCTGC
51 CACTGCTGTG CTGGGGGCCC GGTGCGCCAGG CAAAAGGCC TCCCACGTTT
101 GAGGGGAGTC ATGAGCCGTT TCCTGAATGT GTTAAGAAGT TGGCTGGTTA
151 TGGTGTCCAT CATAGCCATG GGGAAACACG TGCAGAGCTT CCGAGACCAC
201 ACTTTTCTCT ATGAAAAGCT CTACACTGGC AAGCCAAACC TTGTGAATGG
251 CCTCCAAGCT CGGACCTTTG GGATCTGGAC GCTGCTCTCA TCAGTGATCC
301 GCTGCCTCTG TGCCATTGAC ATTCACAACA AGACGCTCTA TCACATCACA
351 CTCTGGACCT TCCTCCTTGC CCTGGGGCAT TTCCTCTCTG AGTTGTTTGT
401 CTATGGAACT GCAGCTCCCA CGATTGGCGT CCTGGCACCC CTGATGGTGG
451 CAAGTTTCTC CATCCTGGGT ATGCTGGTCG GGCTCCGGTA TCTAGAAGTA
501 GAACCAGTAT CCAGACAGAA GAAGAGAAAC TGAGGCCAGC ATTATCACCT
551 CCAGGACTTT CTCGTTTTCC ACCTTGGCCA TCTTCTTCTC TCGTCGTCTC
601 TCCCTTTTAA TTTCTTTTCT ATTCCATCAT CTGCCCTTTT ACTCACTTTT
651 AGCCTCTTTT TTTAATTTT AAAATTTTAA GATATGCATA CTGAAAAGTA
701 TATAACATGT ACGTACAATT TAAAGAATAA TTTTAAAGTG AATACTACGT
751 AACTCCATCC AAGTCAAGAA ATTGCCAGCT TCTCGGAAGC CCACTGTGTC
801 TCCTTCCCCT ACCTGCAACC TCTTCCAGGC TCCCTTTTCC AGCCTTCCCC
851 TTTTCCCCTT TTATTTTCAT GCCTTGATTT GACTTGTGTG GTGGGAACAT
901 GTGAACTATG AAACCTTAAAC CTGTGCCCCA CCCAGAGCAG CTGTGACCAA
951 GGGCTGCCTC AAGGGGTTGT CCACGCAGGT TGGGCTCCTC TCTGCTGCTG
1001 GACCCAAGAC TCTGAACCTT CCAAGGGACA GGCAGTTCTT CTGAGAAGGG
1051 CTCCCCTGTG TGTGAGCAAG ACCACAGCTC TCCTTCTATC TACAGATGCA
1101 TGAGGGTTGG AAGAGTCTGG GCTGTTTTTA GACCTTCTGG TCAGCTGTAT
1151 TTGTGTAACA ACTTTTGTA TAAATAGAAA AACCTCTGCT TCAAAAAAAA
1201 AAAAAAAAAA AA

```

BLAST Results

Entry G38566 from database EMBL:
SHGC-64295 Human Homo sapiens STS genomic, sequence tagged site.
Score = 1398, P = 1.4e-56, identities = 284/288

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 0 bp to 530 bp; peptide length: 177
Category: similarity to unknown protein
Classification: unclassified

```

1 AGAVLGELVC GSGCCCHCCA GGPVARQKAL PRLRGVMSRF LNVLRSLVLM
51 VSIIAMGNLT QSERDHTFLY EKLYTGKPNL VNGLQARTFG IWTLSSVIR
101 CLCAIDIHKN TLYHITLWTF LLAGHFLSE LFVYGTAAPT IGVLA PLMVA

```


BLASTP hits

Alert BLASTP hits for DKFZphfbr2_64h6, frame 3

HSPs :

Pedant information for DKFZphfbr2_64h6, frame 3

Report for DKFZphfbr2_64h6.3

(No Pfam data available for DKFZphfbr2_64h6.3)

DKFZphfbr2_64j18

group: Intracellular transport and trafficking

DKFZphfbr2_624j18.1 encodes a novel 180 amino acid protein nearly identical to the microsomal signal peptidase 23 kd subunit of canis familiaris, gallus gallus and C. elegans.

The new protein is identical to canine and chicken microsomal signal peptidase 23 kd subunit. The canine microsomal signal peptidase is a protein complex comprised of five subunits (25, 22/23, 21, 18, and 12 kDa). The 23kDa subunit is tightly associated with the 18- and 21-kDa subunits, that are integral membrane proteins.

The new protein can find application in modulation of protein transport into microsomal compartments and as a tool for proteomic analysis.

strong similarity to dog signal peptidase (EC 3.4.99.-)

complete cDNA, complete cds, potential start at Bp 109, EST hits,

Sequenced by Qiagen

Locus: unknown

Insert length: 690 bp

Poly A stretch at pos. 666, polyadenylation signal at pos. 646

```

1 GCCGGAACGC GCGCACC GCA GACGGCGCGG ATCGCAGGGA GCCGGTCCGC
51 CGCCGGAACG GGAGCCTGGG TGTGCGTGTG GAGTCCGGAC TCGTGGGAGA
101 CGATCGCGAT GAACACGGTG CTGTCGCGGG CGAACTCACT GTTCGCCTTC
151 TCGCTGAGCG TGATGGCGGC GTCACCTTC GGCTGCTTCA TCACCACCGC
201 CTTCAAAGAC AGGAGCGTCC CGGTGCGGCT GCACGTCTCG CGGATCATGC
251 TAAAAAATGT AGAAGATTTT ACTGGACCTA GAGAAAGAAG TGATCTGGGA
301 TTTATCACAT CTGATATAAC TGCTGATCTA GAGAATATAT TTGATTGGAA
351 TGTTAAGCAG TTGTTTCTTT ATTTATCAGC AGAATATTC AAAAAAATA
401 ATGCTCTGAA CCAAGTTGTC CTATGGGACA AGATTGTTT GAGAGGTGAT
451 AATCCGAAGC TGCTGCTGAA AGATATGAAA ACAAATATT TTTTCTTTGA
501 CGATGGAAAT GGTCTCAAGG GAAACAGGAA TGCTCACTTG ACCCTGTCTT
551 GGAACGTCGT ACCAAATGCT GGAATTCTAC CTCTTGTCAC AGGATCAGGA
601 CACGTATCTG TCCCATTTCC AGATACATAT GAAATAACGA AGAGTTATTA
651 AATTATTCTG AATTGAAAC AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

89034208:

cDNA-derived primary structure of the glycoprotein component of canine microsomal signal peptidase complex.

Peptide information for frame 1

ORF from 109 bp to 648 bp; peptide length: 180
 Category: strong similarity to known protein
 Prosite motifs: TONB_DEPENDENT_REC_1 (1-58)
 RGD (148-151)

```

1 MNTVLSRANS LFAFSLSVMA ALTFGCFITT AFKDRSVVPR LHVSRIMLKN
51 VEDFTGPRER SDLGFIITSDI TADLENIFDW NVKQLFLYLS AEYSTKNNAL
101 NQVVLWDKIV LRGDNPKLLL KDMKTKYFFF DDGNGLKGNR NVTTLISWNV
151 VPNAGILPLV TSGHVSVPFF PDTYEITKSY

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64j18, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64j18, frame 1

Report for DKFZphfbr2_64j18.1

[LENGTH] 180
[MW] 20253.39
[pI] 8.66
[ROMOL] PIR:A31788 signal peptidase (EC 3.4.99.-) (SPC 22/23) - dog 1e-100
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YLR066w]
6e-15
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YLR066w] 6e-15
[PIRKW] transmembrane protein 2e-92
[PIRKW] glycoprotein 2e-92
[PIRKW] hydrolase 2e-92
[PROSITE] RGD 1
[PROSITE] MYRISTYL 2
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha Beta
[KW] SIGNAL_PEPTIDE 32

SEQ. MNTVLSRANSLFAFSLSVMAALTFGCFITTA FKDRSVFVRLHVSRIMLKNVEDFTGPRER
PRD cccccchhhhhhhhhhhhhhhhhhhhhheeeccccceehhhhhhhhhhhhhcccccc

SEQ. SDLGFI TS DITADLENI FDNVVKQLFLYL SAEYSTKNNALNQVVLWDKIVLRGDNPKLLL
PRD cccccchhhhhhhccccchhhhhhhhhhhhhhhhhccccceeeeeeccccchhhhh

SEQ. KDMKTKYFFFDGNGLGKGNRNVTLT LSWNVVPNAGILPLVTGSGHVSVPFPD TYEITKSY
PRD hhccccceeeccccccccccccceeeccccceeeccccceeecccccccccccc

Prosites for DKFZphfbr2_64j18.1

PS00001	141->145	ASN_GLYCOSYLATION	PDOC00001
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	135->141	MYRISTYL	PDOC00008
PS00013	16->27	PROKAR_LIPOPROTEIN	PDOC00013
PS00016	112->115	RGD	PDOC00016
PS00430	1->22	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphfbr2_64j18.1)

DKFZphfbr2_64k24

group: transmembrane proteins

DKFZphfbr2_64k24 encodes a novel 412 amino acid protein with weak similarity to several known proteins.

The novel protein contains 5 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to AMAC1 "testicular condensing enzyme" ;
membrane regions: 5

Summary DKFZphfbr2_64k24 encodes a novel 412 amino acid protein, with
similarity to AMAC1"; product: "testicular condensing enzyme

similarity to AMAC1 "testicular condensing enzyme"

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1958 bp

Poly A stretch at pos. 1939, polyadenylation signal at pos. 1918

```
1 GGGCCCGCCT CGATTTTCCC AGGCGAGGGC ACGCCCGCGT CAGTCGCCTC
51 CGGGGACACCT TCCTCGCCAC GACACGCAGG TAACCGGGCC CCGGGAGCCG
101 GTCGGCGCGG GCGGACTGGG ACCTTGATCC TGCCTGCCCG GCCGCCCGAC
151 AAGGGAATGA GAGCGGACCC CGAACTCCAC ACACCCGCGT TTAGCCGCCA
201 CACCTAAGGG GCAGAACAGT CTTTTTGGGT AAGGGCCGGG CTGGGGGCGA
251 CGCGCCCCCG CCGCTTTGCA GACTTCGGGG TGCTCTGCAC GACGCCTGAA
301 AGGCGCGCGG GCCCGCATTT CTCTGTGCTG CCCTCCTGGA GAACCGGGAC
351 ACGGGGACGG GAGGGCCAGC ATCGGCTACG GCCCGGTTTC CCGTTTCTTT
401 CCTCTGTGCG GTCTGGGCCC TCCTGCAGCG TCCATGATGA AGGCCAGGGG
451 CTGTTGCTTT CCTCTCGCCC AGTAGCCAAC CCAAGCAAGG GAATTAATTA
501 TCTGAAGAAA TGGATACTTC TCCCTCCAGA AAATATCCAG TTAAAAAACG
551 GGTGAAATA CATCCCAACA CAGTGATGGT GAAATATACT TCTCATATC
601 CCCAGCCTGG CGATGATGGA TATGAAGAAA TCAATGAAGG CTATGGGAAT
651 TTTATGGAGG AAAATCCAAA GAAAGGTCTG CTGAGTGAAA TGAAAAAATA
701 AGGGAGAGCT TTCTTTGGAA CCATGGATAC CCTACCTCCA CCAACAGAA
751 ACCCAATGAT CAATGAGATT GGACAATTCC AGAGCTTTGC AGAAAAAAC
801 ATTTTTCAT CCAGAAAAAT GTGGATAGTG CTGTTTGGAT CTGCTTTGGC
851 TCATGGATGT GTAGCTCTTA TCACTAGGCT TGTTCCTGAT CGGTCTAAAG
901 TTCCATCTCT AGAAGTATT TTTATCCGTT CTGTTTTCCT GGTCTTATCT
951 GTGTTAGTTG TGTGTTACTA TCAGGAGGCC CCCTTTGGAC CCAGTGGATA
1001 CAGATTACGA CTCTTCTTTT ATGGTGATAG CAATGTCAAT TCTATCACTT
1051 GTGCTTATAC ATCATTTCCT ATAGTTCCCT CCAGCAATGG GACCACTATG
1101 TGGAGAGCCA CAACTACAGT CTTAGTGCC ATTTTGGCTT TTTTACTCGT
1151 AGATGAGAAA ATGGCTTATG TTGACATGGC TACAGTTGTT TGCAGCATCT
1201 TAGGTGTTTG TCTTGTCATG ATCCCAACA TTGTTGATGA AGACAATTCT
1251 TTGTTAAATG CCTGGAAGA AGCCTTTGGG TACACCATGA CTGTGATGGC
1301 TGGACTGACC ACTGCTCTCT CAATGATAGT ATACAGATCC ATCAAGGAGA
1351 AGATCAGCAT GTGGACTGCG CTGTTTACTT TTGGTTGGAC TGGGACAATT
1401 TGGGGAATAT CTACTATGTT TATTCTTCAA GAACCCATCA TCCCATTAGA
1451 TGGAGAAACC TGGAGTTATC TCATTGCTAT ATGTGTCTGT TCTACTGCAG
1501 CATTCTTAGG AGTTTATTAT GCCTTGGACA AATTCCATCC AGCTTTGGTT
1551 AGCACAGTAC AACATTGGA GATTGTGTA GCTATGGTCT TGCAGCTTCT
1601 CGTGTCTGAC ATATTTCCCT GCATCTATGA TGTTTTTGA GGGGTAATCA
1651 TTATGATTAG TGTTTTTGT CTTGCTGGCT ATAACTTTA CTGGAGGAAT
1701 TTAAGAAGGC AGGACTACCA GGAAATACTA GACTCTCCCA TTAATGAAT
1751 ACCTGATTAT TATTGTCTCA TTAATGTTC GTTATTATA TGTATACTGC
1801 CATTTTAATG TTTACCTATG AATGCTTTT GTGTTATATA ACTGACAGAG
1851 TGCTATAAAA TATATAATAT ATACAAATGC AGAAAAATTA TTCTAGTCTA
1901 ATATATTCAA ATACAAATAT TAAATATATG AAATACGTTA AAAAAAATA
1951 AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 510 bp to 1745 bp; peptide length: 412
 Category: similarity to known protein

```

1 MDTSPSRKYP VKKRVKIHPN TVMVKYTSY PPGDDGYEE INEGYGNFME
51 ENPKKGLLSE MKKKGRAFFG TMDLPPPTTE DPMINEIGQF QSFAEKNIFQ
101 SRKMWIVLFG SALAHGCVAL ITRLVSDRSK VPSLELIFIR SVFQVLSVLV
151 VCYQQEAPFG PSGYRLRLFF YGVCNVSIT CAYTSFSIYP PSNGTTMWRA
201 TTTVFSAILA FLLVDERMAY VDMATVVCSE LGVCLVMIPN IVDEDNSLLN
251 AWKEAFGYTM TVMAGLTAL SMIVYRSIKE KISMWTALFT FGWTGTIWI
301 STMFILQEPI IPLDGETWSY LIAICVCSTA AFLGVYYALD KFHAPALVSTV
351 QHLEIVVAMV LQLLVLFHIFP SIYDVFGGVI IMISVFVLG YKLYWRNLRR
401 QDYQEILDSP IK
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64k24, frame 3

TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds., N = 1, Score = 191, P = 1.9e-12

TREMBL:BMAJ733_6 product: "hypothetical protein"; Bacillus megaterium bgaM gene, N = 1, Score = 137, P = 1.6e-06

PIR:G71841 hypothetical protein jhp1155 - Helicobacter pylori (strain J99), N = 1, Score = 129, P = 1.3e-05

>TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds.

Length = 362

HSPs:

Score = 191 (28.7 bits), Expect = 1.9e-12, P = 1.9e-12
 Identities = 39/105 (37%), Positives = 66/105 (62%)

Query: 289 FTFGWTGTIWGISTMFILQEPIIPLDGETWSYLIAICVCSTAAFLGVYYALDKFHAPALVS 348
 F FG G + + +F+LQ P++P D +WS ++A+ + + +F+ V YA+ K HPALV
 Sbjct: 248 FLFGLVGLMVSVPGLFVLQTPVLPQDTLSWSCVAVGLLALVSFVCVSYAVTKAHPALVC 307

Query: 349 TVQHLEIVVAMVQLLVLFH--IFPSIYDVFGGVIIMISVFVLGAYKL 393
 V H E+VVA++LQ VL+ + PS D+ G +++ S+ ++ L
 Sbjct: 308 AVLHSEVVVALMLQYYVLYETVAPS--DIMGAGVVLGSAIITAQNL 352

Pedant information for DKFZphfbr2_64k24, frame 3

Report for DKFZphfbr2_64k24.3

```

[LENGTH]      412
[MW]           46449.87
[pI]           6.99
[HOMOL]       TREMBL:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus
musculus testicular condensing enzyme (AMAC1) mRNA, complete cds. 8e-14
[PROSITE]     MYRISTYL 6
[PROSITE]     CK2_PHOSPHO_SITE 3
[PROSITE]     PKC_PHOSPHO_SITE 4
[PROSITE]     ASN_GLYCOSYLATION 1
[KW]          TRANSMEMBRANE 5
  
```

SEQ MDTSPSRKYPVKKRVKIHPNTVMV KYTSYPPGDDGYEINEGYGNFMEENPKKGLLSE


```

PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccchhh
MEM      .....

SEQ      MKKKGRAFFGTMDTLPPPTEDPMINEIGQFQSFAEKNIFQSRKMWIVLFGSALAHGCVAL
PRD      hhhhcccccccccccccccccccccccccccccchhhhhhhhhccceccccccccchhhhc
MEM      .....

SEQ      ITRLVSDRSKVPSELEIFIRSVFQVLSVLVVCYYQEAPFGPSGYRLRLFFYGVCNVISIT
PRD      chhhhccccccccchhhhhhhhhhhheeeeecccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMM.....

SEQ      CAYTSFSIVPPSNGTTMWRATTTVFSAILAFLLVDEKMAVVDMAVVCVILGVCLVMIPN
PRD      eccccccccccccccccccccchhhhhhhhhhhhhhhhhheeeeeeeeeeeeeeeeecc
MEM      .....

SEQ      IVDEDNSLLNAWKEAFGYTMTVMAGLTALSMIVYRSIKEKISMWTALFTFGWTGTIWGI
PRD      cccccchhhhhhhhhhhheeeeeehhhhhhhhhccchhhhhhhhhhhcccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMM.....

SEQ      STMFILQEPIIPLDGETWSYLIAICVCSTAAFLGVYALDKFHPALVSTVQHLEIVVAMV
PRD      ceeeeccccccccccccccccchhhhhhhhhccccccccccccchhhhhhhhhhhhhhh
MEM      MMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM

SEQ      LQLLVLFHIFPSIYDVFVGVIIMISVFVLGYKLYWRNLRRQDYQEILDSPK
PRD      hhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhcccccc
MEM      MMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMM.....

```

Prosites for DKFZphfbr2_64k24.3

PS00001	193->197	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00006	92->96	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00008	70->76	MYRISTYL	PDOC00008
PS00008	88->94	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	265->271	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_64k24.3)

DKF2phfbr2_6a17

group: brain derived

DKF2phfbr2_6a17 encodes a novel 100 amino acid protein with very weak similarity to human finger protein zFOC1.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1424 bp

Poly A stretch at pos. 1405, polyadenylation signal at pos. 1389

```
1 GGGACTGAGG GGGTGGGCTT ACTCCCTGGG CAGTCTTGGG GGCCAGAGCT
51 GAGGCCAGTC CATATTACAG TGGCTGGGCT GTTTTTCCTA GTAGCCCTTA
101 GCATTGGCTG GGATTCCCTG TCCTGGGTGC GCCTCCACCT CCCTTCTGAT
151 GCTTCCTGGC TATGGTGGGG TGGGAACCTC AGTTTCCCCC AAAGCTTCTC
201 CTGGATGCTG GCTTCAGGTT GAAGACCTCG GTTCTTCCAG TTCCTCACGG
251 GTTAGGTAGC GGCTCCTGCA TCACCTTCAG AATCAGTTCC AACCCCACT
301 CTCCTTAGGC TTTGTGCTCT GCTCTGCCCT GCCAGGCTGC CCTTGTCCAT
351 GTGAGTAGCA TGGGCGGGTG GTGGGGACGG CAGTGGTGAT GAAGGGGGTG
401 CACCACAGGC CTCATGAAGC AGTTCCCA CA TGGGCGTGTG GCTGGGGCGT
451 GGCCACCACA GAGCACATGG CTGTGTCTAG GCGCAAGCAC TTAGCAGTA
501 TCTGTTTACA TGGCAAGGA TCAAGCCGAC TACCTGTGCT GTCTACTGGG
551 ACAGCAGTCT CCGAGCTACT CCGTACCTCC CTCTGCCAGG TCGTGGAGTT
601 AGGCCCCAGT CCCTACTTGT CACTGGTTCC CACTGTGCTC CTAAGTGTGC
651 AGCACCTGGG AGCTCTGGCC TGGGGCTGGA GGCCCTGGTA GGAGCTGCAG
701 TTGGAGGCCG TTCTGTGCCC AGCAGCGGTG AGCGGCTCCC ATGGGGCCCTG
751 TGTCTGCAGG GAGCCAGGGC TGGCGCACAT GTGCTGTGAA ACTGGCACCC
801 ACCTGGCGTG CTGTGCGCG CACTTGCTTC CTGCAGCACC TCCTACCCCTG
851 CTCCTGTGCC TCCCTCTCCC CGCGCCTGGC TCAGGAGTGC TGGAAAAGCT
901 CACGCTCGG CCTGGGAGCC TGGCCTCTTG ATATACCTCG AGCTTCCCCT
951 GTGCTCCCA GCCCAGGAC CACTGGCCCC TTGGCCTGAG GGGCTGGGGG
1001 CCCCACGACC TGCAGCGTCG AGTCCGGGAG AGAGCCCGGA GCGGCTGGCC
1051 ATCTCGGCTC GGCCTTGCTG AGAGCCTCCG CCCTGGCTTT CTCCCTGTCT
1101 GGTTCAGTGT GCTCAGTTG GTGCTACACA GCTAGAATAG ATATATTAG
1151 AGAGAGAGAT ATTTTAAAG CAAAGCCAC AATTAGCTGT CCTTTAACAC
1201 CGCAGAACCC CCTCCAGAA GAAGAGCGAT CCCTCGGACG GTCCGGGCGG
1251 GCACCCCTCAG CCGGGCTCTT TGCAGAAGCA GCACCGCTGA CTGTGGGCCC
1301 GGCCCTCAGA TGTGTACATA TACGGCTATT TCCTATTTTA CTGTTCTTCA
1351 GATTAGTAC TTGTAATAA ACACACACAT TAAGGAGAGA TTAACATTT
1401 TTGCCAAAAA AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 389 bp to 688 bp; peptide length: 100
Category: putative protein

```
1 MKGVHHRPHE AVPTWACGWG VATTEHMAVS RRRKFSSICL HAQGSSRLPV
51 LSTGTAVSEL LRSLCQVVE LGPSPYLSLV PTVLLTVQHL GALAWGWRPW
```

BLASTP hits

Entry S70007 from database PIR:
finger protein zfOC1 - human (fragment)
Length = 183
Score = 62 (21.8 bits), Expect = 0.24, Sum P(2) = 0.22
Identities = 18/47 (38%), Positives = 24/47 (51%)

Alert BLASTP hits for DKFZphfbr2_6a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6a17, frame 2

Report for DKFZphfbr2_6a17.2

[LENGTH] 100
[MW] 10944.82
[pI] 9.49
[PROSITE] MYRISTYL 2
[PROSITE] PKC_PHOSPHO_SITE 2
[KW] Alpha_Beta

SEQ MKGVHHRPHEAVPTWACGWGVATTEHMAVSRKHFSSICLHAQGSSRLPVLSTGTAVSEL
PRD cccccccccccccccccccccchhhhhhhhhccccccccccccccccccccchhhh

SEQ LRTSLCQVVELGPSPYLSLVPTVLLTVQHLGALAWGWRPW
PRD hhhhheeeeeccccccccccccchhhhhhhhhchhhhhcccc

Prosites for DKFZphfbr2_6a17.2

PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	54->60	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_6a17.2)

DKFZphfbr2_6b24

group: metabolism

DKFZphfkd2_6b24 encodes a novel 334 amino acid protein with similarity to several bacterial dTDP-4-dehydrorhamnose reductases (EC 1.1.1.133).

The novel protein seems to be a human enzyme similar to dTDP-4-dehydrorhamnose reductases. EC 1.1.1.133 catalyses the reaction: dTDP-6-deoxy-L-mannose + NADP(+) \rightleftharpoons dTDP-4-dehydro-6-deoxy-L-mannose + NADPH.

The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

similar to dTDP-6-deoxy-L-mannose-dehydrogenases

complete cDNA, EST hits, complete cds
Nucleotide sugars metabolism seems to be a dehydrogenase
localisation: region of primer A missing

Sequenced by AGOWA

Locus: /map="5"

Insert length: 2054 bp

Poly A stretch at pos. 2028, polyadenylation signal at pos. 2015

```

1  GGGGGAGGCC CGCGTCGATC CTGGGTGGGA GGAGGTGGCG GCCGCTGAGG
51 CTGCGGCGTG AAGACGGCGG GCATGGTGGG GCGGGAGAAA GAGCTCTCTA
101 TACACTTTGT TCCCGGGAGC TGTGCGCTGG TGGAGGAGGA AGTTAACATC
151 CCTAATAGGA GGGTTCTGCT TACTCGTGCC ACTGGGCTTC TTGGCAGAGC
201 TGTACACAAA GAATTTTCAGC AGAATAATTG GCATGCAGTT GGCTGTGGTT
251 TCAGAAAGAGC AAGACCAAAA TTTGAACAGG TTAATCTGTT GGATTCTAAT
301 GCAGTTCATC ACATCATTCA TGATTTTCAG CCCCATGTTA TAGTACATTG
351 TGCAGCAGAG AGAAGACCAG ATGTTGTAGA AAATCAGCCA GATGCTGCCT
401 CTCAACTTAA TGTGGATGCT TCTGGGAATT TAGCAAAGGA AGCAGCTGCT
451 GTTGGAGCAT TTCTCATCTA CATTAGCTCA GATTATGTAT TTGATGGAAC
501 AAATCCACCT TACAGAGAGG AAGACATACC AGCTCCCCTA AATTTGTATG
551 GCAAAACAAA ATTAGATGGA GAAAAGGCTG TCCTGGAGAA CAATCTAGGA
601 GCTGCTGTTT TGAGGATTCC TATTCTGTAT GGGGAAGTTG AAAAGCTCGA
651 AGAAAGTGCA GTGACTGTTA TGTTCGATAA AGTGCAAGTT AGCAACAAGT
701 CAGCAACATC GGATCACTGG CAGCAGAGGT TCCCCACACA TGTCAAAGAT
751 GTGGCCACTG TGTGCCGGCA GCTAGCAGAG AAGAGAATGC TGGATCCATC
801 AATTAAGGGA ACCTTTCAC TGTCTGGCAA TGAACAGATG ACTAAGTATG
851 AAATGGCATG TGCAATTGCA GATGCCTTCA ACCTCCCCAG CAGTCACTTA
901 AGACCTATTA CTGACAGCCC TGTCCTAGGA GCACAACGTC CGAGAAATGC
951 TCAGCTTGAC TGCTCCAAAT TGGAGACCTT GGGCATTGGC CAACGAACAC
1001 CATTTCGAAT TGGAAACAAA GAATCACTTT GGCCCTTCCT CATTGACAAG
1051 AGATGGAGAC AAACGGTCTT TCATTAGTTT ATTTGTGTTG GGTTCCTTTT
1101 TTTTAAATTA GAAAAGTATA GTATGTGGCC CTTTTTAAAG AACAAAGGAA
1151 ATAGTTTGTG ATGAGTACTT TAATTGTGAC TCTTAGGATC TTTCAGGTAA
1201 ATGATGCTCT TGCACACTAG AAATTGTCTA AAGAAACTAA AGGGCAGTCA
1251 TGCCCTGTTT GCAGTAATTT TTCTTTTATC CATTATGTTT GTCCTGGCTA
1301 AACTTGGAGT TTGAGTATAG TAAATTATGA TCCTTAAATA TTTGAGGGTC
1351 AGGATGAAGC AGATCTGCTG TAGACTTTTC AGATGAAAT GTTCATTCTC
1401 GTAACCTCCA TATTTTCAGG ATTTTGAAG CTGTTGACCA TTTCATGTTG
1451 ATTATTTTAA ATTGTGTGGA ATAGTATAAA AATCATTGGT GTTCATTATT
1501 TGCTTTGCCT GAGCTCAGAT CAAAATGTTT GAAGAAAGGA ACTTTATTTT
1551 TGCAAGTTAC GTACAGTTT TATGCTTGAG ATATTCAAC ATGTTATGTA
1601 TATTGGAAC TCTACAGCTT GATGCCTCCT GCTTTTATAG CAGTTTATGG
1651 GGAGCACTTG AAAGAGCGTG TGTACATGTA TTTTCTTCT AGGCAACAT
1701 TGAATGCAAA CGTGTATTTT TTTAATATAA ATATATAACT GTCCTTTTCA
1751 TCCCATGTTG CCGCTAAGTG ATATTCATA TGTGTGGTTA TACTCATAAT
1801 AATGGGCCCT GTAAGTCTTT TCACCATTC TGAATAATAA TAAATATGTA
1851 CTGCTGGCAT GTAATGCTTA GTTTCTTGT ATTTACTTCT TTTTTTAAA
1901 TGTAAGGACC AAACCTCTAA ACTAATGTT CTTTGTGTC TTTAATTTT
1951 AAAAATTACA TTCTTCTGAT GTAACATGTG ATACATACAA AAGAATATAG
2001 TTTAATATGT ATTGAAATAA AACACAATAA AATTAAAAAA AAAAAAATAA
2051 AAAA

```

BLAST Results

Entry G37115 from database EMBL:
SHGC-56899 Human Homo sapiens STS genomic.
Score = 446, P = 4.6e-14, identities = 90/91

Medline entries

99109950:
The metabolism of 6-deoxyhexoses in bacterial and animal cells.

Peptide information for frame 1

ORF from 73 bp to 1074 bp; peptide length: 334
Category: similarity to known protein

```

1  MVGREKELSI HFVPGSCLRV EEEVNIPNRR VLVGTATGGL GRAVHKEFQQ
51 NNWHAVGCGF RRARPKFEQV NLLDSNAVHH IIHDFQPHVI VHCAAERRPD
101 VVENQPDAA S QLNVDASGNL AKEAAAVGAF LIYISSDYVF DGTNPPYREE
151 DIPAPLNLYG KTKLDGEKAV LENNLGA AVL RIPILYGEVE KLEESAVTVM
201 FDKVQFSNKS ANMDHWQQR F PTHVKDVATV CRQLAEKRML DPSIKGTFHW
251 SGNEQMTKYE MACAIADAFN LPSSHLRPIT DSPVLGAQRP RNAQLDCSKL
301 ETLGIGQRT P FRIGIKESLW PFLIDKRWRQ TVFH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6b24, frame 1

PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans, N = 1, Score = 293, P = 6.4e-26

TREMBL:SSU51197_21 gene: "rhd"; product:
"dTDP-6-deoxy-L-mannose-dehydrogenase"; Sphingomonas S88 sphingan
polysaccharide synthesis (spsG), (spsS), (spsR), glycosyl transferase
(spsQ), (spsI), glycosyl transferase (spsK), glycosyl transferase
(spsL), (spsJ), (spsF), (spsC), (spsE), Urf 32, Urf 26,
ATP-binding cassette trans., N = 1, Score = 291, P = 1e-25

SWISSPROT:RFB_DHISN PROBABLE DTDP-4-DEHYDRORHAMNOSE REDUCTASE (EC
1.1.1.133) (DTDP-4-KETO- L-RHAMNOSE REDUCTASE) (DTDP-6-DEOXY-L-MANNOSE
DEHYDROGENASE) (DTDP-L- RHAMNOSE SYNTHETASE)., N = 1, Score = 283, P =
7.4e-25

>PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans
Length = 294

HSPs:

Score = 293 (44.0 bits), Expect = 6.4e-26, P = 6.4e-26
Identities = 89/276 (32%), Positives = 151/276 (54%)

```

Query:   30 RVLVTGATGLGRAVHKEFQQNNWHAVGCGFRARPKFEQV NLLDSNAVHHIIHDFQPHV 89
          R+L+TGA G LGR++ K   N + V           F +++++ + + V II F+P+V
Sbjct:   3  RLLITGAGGQLGRSLAKLLVDNGRYEV-----LALDFSELDITNKMVFESIIDSFKPNV 56

Query:   90 IVHCAAERRPDVVENQPDAA S QLNVDASGNLAKEAAAVGAFLIYISSDYVFDG-TNPPYR 148
          I++ AA   D E + +A +NV   LA+ A   + ++++S+DYVFDG + Y+
Sbjct:   57 IINAAAYTSVDQAELEVSSAYSVNVRGVQYLAEAAIRHNSAILHVSTDYVFDGYKSGKYK 116

Query:   149 EEDIPAPLNLYGKTKLDGEKAVLENNLGA AVL RIPILYGEVEKLEESAVTVMFDKVQFSN 208
          E DI  PL +YGK+K +GE+ +L + + +LR   +GE   + V M ++ +
Sbjct:   117 ETDIIHPLCVYGKSKAEGERLLLTSPKSIILRTSWTFGEYGN---NFVKTML-RLAKNR 172

Query:   209 KSANMDHWQQRFP THVKDVATVCRQLAEKRMLDPSIK-GTFHWSGNEQMTKYEMACAIAD 267
          +   Q   PT+ D+A+V Q+AEK ++ ++K G +H++G ++ Y+ A AI D
Sbjct:   173 DILGVVADQIGGPTYSGDIASVLIQIAEKIIVGETVKYGIYHFTGEPVSWYDFAIAIFD 232

Query:   268 AF-----NLPSSHLRPITDSPVLGAQRP RNAQLDCSKLE-TLGI 305
          N+P +   D P L A+RP N+ LD +K++   GI
Sbjct:   233 EAVAQKVLNVPLVNAITTADYPTL-AKRPANSCLDLTKIQQAFGI 277

```


Pedant information for DKFZphfbr2_6b24, frame 1

Report for DKFZphfbr2_6b24.1

```
[LENGTH] 334
[MW] 37551.98
[pI] 6.90
[OMOL] PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans 6e-25
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YGL001c]
6e-04
[EC] 1.1.1.133 dTDP-4-dehydrorhamnose reductase 2e-16
[PIRKW] lipopolysaccharide biosynthesis 2e-16
[PIRKW] NADP 2e-16
[PIRKW] oxidoreductase 2e-16
[PIRKW] streptomycin biosynthesis 1e-19
[SUPFAM] dTDP-dihydrostreptose synthase 1e-20
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 3
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha Beta
```

SEQ	MVGREKELSTHFVPGSCLRVEEVNI PNRRVLVTGATGLLGRAVHKFEQQNNHWAHVCGGF
PRD	ccccceeeccccccccceeeccccccccceeeccccchhhhhhhhhhhccccceeeccc
SEQ	RRARPKEQVNLDSNAVHHIHDHFQPHVIVHCAAERRPDVVENQPDAASQLNVDASGNL
PRD	ccccccccccccchhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhhhhhhhccccchh
SEQ	AKEAAAVGAFLIYISSDYVFDGTNPFPYREEDIPAPLNLYGKTKLDGEKAVLENNLGAAVL
PRD	hhhhhhhhhheeeeeeccccccccccccccccccccccccccccchhhhhhhhhccccceeee
SEQ	RIPILYGEVERKEESAVTVMFDKQVFSNKSANMDHWQQRFPTHVKDVATVCRQLAEKRML
PRD	eeeeeeccccccccchhhhhhhhhhhhhccccceeeccccccccccccchhhhhhhhhhhhhh
SEQ	DPSIKGTFHWGSGNEQMTKYEMACAIADAFNLPSSHLRPITDSPVLGAQRPRNAQLDCSKL
PRD	ccccceeeccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccchhhhh
SEQ	ETLGIGQRTPFRIGIKESLWPFLLDKRWQTVFH
PRD	hhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

Prosites for DKF2phfbr2_6b24.1

PS000001	208->212	ASN_GLYCOSYLATION	PDOC000001
PS000005	16->19	PKC_PHOSPHO_SITE	PDOC000005
PS000005	207->210	PKC_PHOSPHO_SITE	PDOC000005
PS000005	243->246	PKC_PHOSPHO_SITE	PDOC000005
PS000006	162->166	CK2_PHOSPHO_SITE	PDOC000006
PS000006	251->255	CK2_PHOSPHO_SITE	PDOC000006
PS000006	257->261	CK2_PHOSPHO_SITE	PDOC000006
PS000006	298->302	CK2_PHOSPHO_SITE	PDOC000006
PS000008	314->320	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2_6b24.1)

DKFZphfbr2_6i20

group: brain derived

DKFZphfbr2_6i20 encodes a novel 296 amino acid protein with similarity to ribosomal protein L15 precursor of *S. cerevisiae* mitochondria.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ribosomal protein L15 precursor, mitochondrial

complete cDNA, complete cds, EST hits
potential mitochondrial L15 ribosomal protein

Sequenced by AGOWA

Locus: /map="377.5 cR from top of Chr8 linkage group"

Insert length: 1122 bp

Poly A stretch at pos. 1099, polyadenylation signal at pos. 1071

```

1 GGGGGCCCTT GAAAGTTCTT GGATCTGCGG GTTATGGCCG GTCCCTTGCA
51 GGGCGGTGGG GCCCGGGCCC TGGACCTACT CCGGGGCCGT CCGCGTGTGA
101 GCCTGGCCAA CTTAAAGCCG AATCCCGGCT CCAAGAAACC GGAGAGAAGA
151 CCAAGAGGTC GGAGAAGAGG TAGAAAATGT GGCAGAGGCC ATAAAGGAGA
201 AAGGCAAAGA GGAACCCGGC CCCGCTTGGG CTTTGAGGGA GGCCAGACTC
251 CATTTTACAT CCGAATCCCA AAATACGGGT TTAACGAAGG ACATAGTTTC
301 AGACGCCAGT ATAAGCCTAT GAGTCTCAAT AGACTGCAGT ATCTTATTGA
351 TTTGGGTCGT GTTGATCCTA GTCAACCTAT TGACTTAACC CAGCTTGTC
401 ATGGGAGAGG TGTGACCATC CAGCCACTTA AAAGGGATTA TGATGTCCAG
451 CTGTTTGAGG AGGGTGCTGA CACCTTTACG GCAAAAGTTA ATATTGAAGT
501 ACAGTTGGCT TCAGAACTAG CTATGTCTGC CATTGAAAAA AATGGTGGTG
551 TTGTACTAC AGCCTTCTAT GATCCAAGAA GTCTGGACAT TGTATGCAAA
601 CCTGTTCCAT TCTTTCTTCG TGGACAACCC ATTCCAAAAA GAATGCTTCC
651 ACCAGAAGAA CTGGTACCAT ATTACACTGA TGCAAGAAGC CGTGGGTACC
701 TGGCGGATCC TGCCAAATTT CCTGAAGCAC GACTTGAACG CGCCAGGAAG
751 TATGTTTATA TCTTACCTGA TATCACTAAA GATGAACCTC TCAAAATGCT
801 CTGTACTAGG AAGGATCCAA GGCAGATTTT CTTTGGTCTT GCTCCAGGAT
851 GGGTGGTGAA TATGGCCGAT AAGAAAAATCC TAAACCTAC AGATGAAAT
901 CTCCTTAAGT ATTATACCTC ATGAATTCCC GTCCAAGGAA GCAGAGTTGT
951 TAAAGAGTAC TGGAATAGGG GCTGAAGGAT CTATATTCCC TTATTGCATT
1001 TTCCTTATGT ATAATTTTCC AGATGGTGAT GTTACTTTTC ACTGTACTCA
1051 TATGTCTCAT TTTCATCTAA AATTAAATGG CAGGAAACAA GGACTGCATA
1101 GAGAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS500354 from database EMBL:
human STS WI-12392.
Length = 426
Minus Strand HSPs:
Score = 1791 (268.7 bits), Expect = 1.1e-74, P = 1.1e-74
Identities = 375/384 (97%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 34 bp to 921 bp; peptide length: 296
Category: strong similarity to known protein

1 MAGPLQGGGA RALDLLRGLP RVSLANLKPN PGSKKPERRP RGRRRGRKCG

51 RGHKGERQRG TRPRLGFEGG QTPFYIRIPK YGFNEGHSFR RQYKPMNLNR
 101 LQYLIDLGRV DPSQPIDLTQ LVNDRGVTIQ PLKRDYDVQL VEEGADTFTA
 151 KVNIEVQLAS ELAIAAIEKN GGVVTTAFYD PRSLDIVCKP VPFFLRGQPI
 201 PKRMLPPEEL VPYYTDAKNR GYLADPAKFP EARLELARKY GYLDPDITKD
 251 ELFKMLCTRK DPRQIFFGLA PGWVVMADK KILKPTDENL LKYYTS

BLASTP hits

Entry S63258 from database PIR:
 ribosomal protein L15 precursor, mitochondrial - yeast (*Saccharomyces cerevisiae*)
 Length = 322
 Score = 259 (91.2 bits), Expect = 2.0e-22, P = 2.0e-22
 Identities = 71/200 (35%), Positives = 106/200 (53%)

Entry H70161 from database PIR:
 ribosomal protein L15 (rplO) - Lyme disease spirochete
 Length = 145
 Score = 173 (60.9 bits), Expect = 4.8e-13, P = 4.8e-13
 Identities = 45/140 (32%), Positives = 73/140 (52%)

Alert BLASTP hits for DKFZphfbr2_6i20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6i20, frame 1

Report for DKFZphfbr2_6i20.1

[LENGTH] 296
 [MW] 33495.98
 [pI] 9.98
 [HOMOL] TREMBL:AF067212_1 gene: "F37F2.1"; *Caenorhabditis elegans* cosmid F37F2. 1e-38

[FUNCAT] 05.01 ribosomal proteins (*S. cerevisiae*, YNL284c] 7e-15
 [FUNCAT] 30.16 mitochondrial organization (*S. cerevisiae*, YNL284c] 7e-15
 [FUNCAT] j mrna translation and ribosome biogenesis (*M. genitalium*, MG169] 1e-06
 [BLOCKS] BL00475D
 [BLOCKS] BL00475B Ribosomal protein L15 proteins
 [PIRKW] ribosome 2e-13
 [PIRKW] mitochondrion 2e-13
 [PIRKW] protein biosynthesis 2e-13
 [SUPFAM] Escherichia coli ribosomal protein L15 4e-06
 [PROSITE] MYRISTYL 3
 [PROSITE] AMIDATION 2
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 4
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 12.50 %

SEQ MAGPLQGGGARALDLLRGLPRVSLANLKPNGSKKPERRRPRGRRGRKCGRGHKGERQRG
 SEGXXX
 PRD ccc

SEQ TRPRLGFEGGQTPFYIRIPKYGFNEGHSFRQYKPMNLRLQYLIDLGRVDPSQPIDLTQ
 SEG
 PRD ccc

SEQ LVNDRGVTIQPLKRDYDVQLVEEGADTFTAKVNIEVQLASELAIAAIEKNGGVVTTAFYD
 SEG
 PRD ecc

SEQ PRSLDIVCKPVPFFLRGQPIPKRMLPPEELVPYYTDAKNRGYLADPAKFPEARLELARKY
 SEG
 PRD ccc

SEQ GYLDPDITKDELFKMLCTRKDPRQIFFGLAPGWVVMADKKILKPTDENLLKYYTS
 SEG
 PRD ccc

Prosite for DKFZphfbr2_6i20.1

PS00005 33->36 PKC_PHOSPHO_SITE PDOC00005
 PS00005 88->91 PKC_PHOSPHO_SITE PDOC00005

PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	171->177	MYRISTYL	PDOC00008
PS00008	268->274	MYRISTYL	PDOC00008
PS00009	41->45	AMIDATION	PDOC00009
PS00009	45->49	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_6i20.1)

DKFZphfbr2_6ol7

group: nucleic acid management

DKFZphfbr2_6ol7 encodes a novel 455 amino acid protein with strong similarity to DEAD-box ATP-dependent RNA helicases YHR065c and T26G10.1.

The *S. cerevisiae* protein YHR065c is required for maturation of the 35S RNA primary transcript.

The new protein can find application in modulating rRNA maturation.

strong similar to RNA helicases

complete cDNA, complete cds, EST hits
probable start at Bp 27 matches kozak consensus ANNatgG
involved in maturation of r-RNA ??
YHR065c/Rxp3p is involved in maturation of the 35S primary transcript
Drs1p cold-sensitive mutation has slow 27S to 25S pre-rRNA
conversion and is deficient in 60S ribosomal subunits

Sequenced by AGOWA

Locus: unknown

Insert length: 1840 bp

Poly A stretch at pos. 1815, polyadenylation signal at pos. 1793

```
1  GGGGACTTCC  GGAGACCTCA  CACAAGATGG  CGGCACCCGA  GGAACACGAT
51  TCTCCGACCG  AAGCGTCCCA  GCCGATTGTG  GAAGAGGAGG  AACTAAAAC
101  ATTTAAAGAC  CTGGGTGTGA  CAGATGTGTT  GTGTGAAGCT  TGTGACCACT
151  TGGGATGGAC  AAAACCCACC  AAGATTCAGA  TTGAAGCTAT  TCCTTTGGCC
201  TTACAAGGTC  GTGATATCAT  TGGGCTTGCA  GAAACTGGCT  CTGGAAGAC
251  AGGCGCCTTT  GCTTTGCCCA  TTCTAAACGC  ACTGCTGGAG  ACCCCGCAGC
301  GTTTGTTTGC  CCTAGTTCTT  ACCCGGACTC  GGGAGCTGGC  CTTTCAGATC
351  TCAGAGCAGT  TTGAAGCCCT  GGGGTCTCT  ATTGGAGTGC  AGAGTGCTGT
401  GATTGTAGGT  GGAATTGATT  CAATGTCTCA  ATCTTTGGCC  CTTGCAAAAA
451  AACCACATAT  AATAATAGCA  ACTCCTGGTC  GACTGATTGA  CCACCTGGAA
501  AATACGAAAG  GTTCAACTT  GAGAGCTCTC  AAATACTTGG  TCATGGATGA
551  AGCCGACCGA  ATACTGAATA  TGGATTTTGA  GACAGAGGTT  GACAAGATCC
601  TCAAGTGAT  TCCTCGAGAT  CGGAAAACAT  TCCTCTTCTC  TGCCACCATG
651  ACCAAGAAGG  TTCAAAAAC  TCAGCGAGCA  GCTCTGAAGA  ATCCTGTGAA
701  ATGTGCCGTT  TCCTCTAAAT  ACCAGACAGT  TGAAAAATTA  CAGCAATATT
751  ATATTTTAT  TCCCTCTAAA  TTCAAGGATA  CCTACCTGGT  TTATATTCTA
801  AATGAATTGG  CTGGAACTC  CTTTATGATA  TTCTGCAGCA  CCTGTAATAA
851  TACCCAGAGA  ACAGCTTTGC  TACTGCGAAA  TCTTGGCTTC  ACTGCCATCC
901  CCTCCATGG  ACAAATGAGT  CAGAGTAAGC  GCCTAGGATC  CCTTAATAAG
951  TTTAAGGCCA  AGGCCCGTTC  CATTCTTCTA  GCAACTGACG  TTGCCAGCCG
1001  AGGTTTGGAC  ATACCTCATG  TAGATGTGGT  TGTCAACTTT  GACATTCTTA
1051  CCCATTCCAA  GGATTACATC  CATCGAGTAG  GTCGAACAGC  TAGAGCTGGG
1101  CGCTCCGGAA  AGGCTATTAC  TTTTGTCA  CAGTATGATG  TGGAACTCTT
1151  CCAGCGCATA  GAACACTTAA  TTGGGAAGAA  ACTACCAGGT  TTTCCAACAC
1201  AGGATGATGA  GGTATGATG  CTGACAGAAC  GCGTCGCTGA  AGCCCAAAGG
1251  TTTGCCCGAA  TGGAGTTAAG  GGAGCATGGA  GAAAGAAGA  AACGCTCGCG
1301  AGAGGATGCT  GGAGATAATG  ATGACACAGA  GGGTGCTATT  GGTGTCAGGA
1351  ACAAGGTGGC  TGGAGGAAAA  ATGAAGAAGC  GGAAGGCCG  TTAATCACTT
1401  TTATGAAGGC  TCGAGTTCTG  CTGTTCTGTA  AAAGAAATTT  GGAGAATGAA
1451  ACCTGCTCCA  ACAGAGATCA  TGAGACTGAA  ATTGGTCAGA  ATTGTGTCCA
1501  GAATGTGCTC  AGCTAATTCA  GTATTCTTCC  CCATTCTGGG  TTGGAGTTTA
1551  CTGCAGAGTA  ATTCTTACAG  TGCTGATGTC  AAGACTGTTA  CTGTTCTTCG
1601  ACTTTGATT  CTTGCTCATG  ACATGAGTAG  GGTGTGCTCT  TCTGTCACTT
1651  CACACAGACC  TTTTGCTTTT  TTTAGCTGCA  AGTCAAGGAC  TAGGTTGATG
1701  ATGCCCCATG  CCTGTAATTG  TAAAGAAGCT  TGGACATCTG  CAAATGATAT
1751  TTAACCATC  TTGGCTTGTG  CTTTATTCAA  ACTAATGTGA  AACAATAAAT
1801  TTAATATTA  TTTTAAAG  AAAAAAAAAA  AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 1391 bp; peptide length: 455
 Category: strong similarity to known protein

```

1 MAAPEEHDSP TEASQPIVEE EETKTFKDLG VTDVLCACD QLGWTKPTKI
51 QIEAIPALQ GRDIIGLAET GSGKTGAFAL PILNALLETQ QRLFALVLTQ
101 TRELAFOISE QFEALGSSIG VQSAVIVGGI DSMSQSLALA KKPHEIATP
151 GRIDHLENT KGFNLRLKY LVMDEADRIL NMDFETEVDK ILKVIPDRK
201 TFLFSATMTK KVQKLQRAAL KNPVKCAVSS KYQTVEKLQ YIFIPSKFK
251 DTYLVYILNE LAGNSFMIFC STCNNTQRTA LLRLNLGFTA IPLHGQMSQS
301 KRLGSLNKF AKARSILLAT DVASRGDIP HVDVVVNFDI PTHSKDYIHR
351 VGR TARAGRS GKAITFVTQY DVELFQRIE LIGKKLPGFQ TQDDEVMMLT
401 ERVAEAQRFA RMELREHGEK KRSREDAGD NDDTEGAIGV RNKVAGGKMK
451 KRKGR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6ol7, frame 3

PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*, N = 1, Score = 1497, P = 1.6e-153

PIR:S46713 hypothetical protein YHR065c - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 1154, P = 3.6e-117

TREMBL:ATH010462_1 gene: "RH10"; product: "RNA helicase"; *Arabidopsis thaliana* mRNA for DEAD box RNA helicase, RH10, N = 1, Score = 1122, P = 8.9e-114

TREMBL:AC002985_2 product: "R27090_2"; Human DNA from chromosome 19-specific cosmid R27090, genomic sequence, complete sequence., N = 1, Score = 950, P = 1.5e-95

>PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*
 Length = 489

HSPs:

Score = 1497 (224.6 bits), Expect = 1.6e-153, P = 1.6e-153
 Identities = 283/442 (64%), Positives = 364/442 (82%)

```

Query:   19 EEEETKTFKDLGVTDVLCACDQLGWTKPTKIQIEAIPALQGRDIIGLAETGSGKTGAF 78
          E+ + K+F +LGV+ LC+AC +LGW KP+KIQ A+P ALQG+D+IGLAETGSGKTGAF
Sbjct:   39 EDVKEKSFALGVSQLCDACQRLGWMKPSKIQQAALPHALQKDVIGLAETGSGKTGAF 98

Query:   79 ALPILNALLETQQLFALVLTPTRELAFOISEQFEALGSSIGVQSAVIVGGIDSMSQSLA 138
          A+P+L +LL+ PQ F LVLTPTRELAFOI +QFEALGS IG+ +AVIVGG+D +Q++A
Sbjct:   99 AIPVLQSLLDHPQAFCLVLTPTRELAFOIQGQFEALGSGIGLIAAVIVGGVDMAAQAMA 158

Query:   139 LAKKPHIIATPGRLIDHLENTKGFNLRLKYLVMDEADRILNMDFETEVDKILKVIPRD 198
          LA++PHII+ATPGRL+DHLENTKGFNL+ALK+L+MDEADRILNMDFE E+DKILKVIPR+
Sbjct:   159 LARRPHIIVATPGRLVDHLENTKGFNLKALKFLIMDEADRILNMDFEVELDKILKVIPRE 218

Query:   199 RKTFLSATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQYYIFIPSKFKDITYLVYIL 258
          R+T+LFSATMTKKV KL+RA+L++P + +VSS+Y+TV+ L+Q+YIF+P+K+K+TYLVY+L
Sbjct:   219 RRTYLSATMTKKVSKLERASLRDPARVSVSSRYKTVDNLKQHYIFVPNKYKETYLVYLL 278

Query:   259 NELAGNSFMIFCSTCNNTQRTALLRLNLGFTAIPHLHGQMSQSKRLGSLNKFKAARSILL 318
          NE AGNS ++FC+TC T + A++LR LG A+PLHGQMSQ KRLGSLNKF+KAR IL+
Sbjct:   279 NEHAGNSAIVFCATCATTMQIAVMLRQLGMQAVPLHGQMSQEKRLGSLNKFKSKAREILV 338

Query:   319 ATDVASRGDIPHDVVVNFDIPTHSKDYIHRVGR TARAGRSKAITFVTQYDVELFQRI 378
          TDVA+RGLDIPHVD+V+N+D+P+ SKDY+HRVGR TARAGRS AIT VTQYDVE +Q+I
Sbjct:   339 CTDVAARGLDIPHVMVINYDMPQSQSKDYVHRVGR TARAGRSIAITVVTQYDVEAYQKI 398

Query:   379 EHLIGKKLPGFPTQDDEVMMILTERVAEAQRFA RMELREHGEKKK-----RSREDAGDND 433
          E +GKKL + ++EVM+L ER EA AR+E++E EKKK R +D GD ++
Sbjct:   399 EANLGKKLDEYKCVENEVMVLVERTQEATENARIEMKEMDEKKKSGKKRRQNDDFGDTTE 458

Query:   434 TEGAIGVRNKVAGGKMKRRKGR 455

```


+ G + K GG+ GR
 Sbjct: 459 SGGRFKMGIKSMGGRGGSGGGR 480

Pedant information for DKFZphfbr2_6ol7, frame 3

Report for DKFZphfbr2_6ol7.3

[LENGTH] 455
 [MW] 50646.80
 [pI] 9.18
 [HOMOL] PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*
 1e-167
 [FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YHR169w] 2e-79
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-71
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-66
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-63
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 1e-58
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YDL084w] 1e-55
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YOR204w] 5e-55
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 5e-55
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H.
 influenzae, HI0892] 9e-48
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 2e-45
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-42
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 7e-16
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 5e-06
 [BLOCKS] BL00175B Phosphoglycerate mutase family phosphohistidine proteins
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-60
 [PIRKW] RNA binding 7e-69
 [PIRKW] DEAD box 7e-69
 [PIRKW] transmembrane protein 9e-41
 [PIRKW] DNA binding 3e-55
 [PIRKW] recF recombination pathway 3e-11
 [PIRKW] ATP 1e-126
 [PIRKW] purine nucleotide binding 7e-69
 [PIRKW] P-loop 1e-126
 [PIRKW] hydrolase 1e-55
 [PIRKW] protein biosynthesis 7e-69
 [PIRKW] ATP binding 3e-61
 [SUPFAM] ATP-dependent RNA helicase eIF-4A 8e-06
 [SUPFAM] WW repeat homology 4e-58
 [SUPFAM] translation initiation factor eIF-4A 7e-69
 [SUPFAM] DEAD/H box helicase homology 1e-126
 [SUPFAM] recQ helicase homology 5e-12
 [SUPFAM] ATP-dependent RNA helicase homology 8e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-126
 [SUPFAM] ATP-dependent RNA helicase DBP1 4e-60
 [SUPFAM] ATP-dependent RNA helicase DHM1 1e-58
 [SUPFAM] recQ protein 3e-11
 [SUPFAM] tobacco ATP-dependent RNA helicase DB10 4e-58
 [SUPFAM] Bloom's syndrome helicase 5e-12
 [PROSITE] DEAD ATP HELICASE 1
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta

SEQ MAAPEEHDSPTEASQPIVEEEETKTFKDLGVTDLVCEACDQLGWTKPTKIQIEAIPALQ
 PRD cccccccccccccchhhhhhhhhccccchhhhhhhhhcccccccccccccccccc
 SEQ GRDIIGLAETGSGKTGAFALPILNALLETPQRLFALVLTPTRELAFQISEQFEALGSSIG
 PRD ccceeeccccccccchhhhhhhhhccccceeeccccchhhhhhhhhhhhhhhhhcc


```

SEQ  VQSAVIVGGIDSMQSLALAKKPHIIATPGRLIDHLENTKGFNLRAKYLVMDEADRIL
PRD  eeeeeeeccchhhhhhhhhccceeeeeccccccccccccccccccccceehhhhhhhh

SEQ  NMDFETEVDKILKVIPRDRKTLFSAATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQ
PRD  hhcchhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhccceeeeeccccchhhh

SEQ  YYIFIPSKFKDTYLVYILNELAGNSFMIFCSTCNNTQRTALLRLNGFTAIPLHGQMSQS
PRD  hhhhhhhhhhhhhhhhhhhhhhhccceeeeeccchhhhhhhhhhhhhhhccceccccchhh

SEQ  KRLGSLNKFKAARSILLATDVASRGLDIPHDVVVNFDIPTHSKDYIHRVGR TARAGRS
PRD  hhhhhhhhhhhhhhhhhhhhhhhccceeeeecccccccccccccccccccccccccc

SEQ  GKAITFVTQYDVELFQRIEHLIGKKLPGFPTQDDEVMLTERVAEAQR FARMELREHGEK
PRD  cceeeeeccchhhhhhhhhhhhhhhccceccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KKRSREDAGDNDDEGAIGVRNKVAGGKMKRKRGR
PRD  hhhcccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_6ol7.3

PS00001	274->278	ASN_GLYCOSYLATION	PDOC00001
PS00004	421->425	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	229->232	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	300->303	PKC_PHOSPHO_SITE	PDOC00005
PS00005	354->357	PKC_PHOSPHO_SITE	PDOC00005
PS00005	360->363	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	25->29	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006
PS00006	424->428	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	128->134	MYRISTYL	PDOC00008
PS00009	382->386	AMIDATION	PDOC00009
PS00017	68->76	ATP_GTP_A	PDOC00017
PS00039	172->181	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphfbr2_6ol7.3

HMM_NAME DEAD and DEAH box helicases

HMM	*gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAG		
Query	30	GVTDLVCEACDQLGWTkPTKIQIEAIPALQGRDIIGLAETGSGKTGAF	78
HMM	lIPMLQHIDwdPwpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMngIR		
Query	79	ALPILNALLETp----QR-LFALVLTPTRELAfQISEQFEALGSSIG-VQ	122
HMM	ImcIYGGtnMRdQMRmLeRGpPHIVATPGRLIDHIER.gtlDLDrIeML		
Query	123	SAVIVGGIDSMQSLALAKKPHIIATPGRLIDHLENTKGFNLRAKYL	171
HMM	VMDEADRMLDMGFIDQIRrImrqIPmpwnRQTMMFSATMPdeIqELARrF		
Query	172	VMDEADRILNMDFETEVDKILKVIP--RDRKTLFSAATMTKKVQKLQRAA	219
HMM	MRNPiRInIdMdElTtnEnIkQwYiyVerEMWkfdeLcrLIe*		
Query	220	LKNPVKCAVSSKYQTVE-KLQQYYIFIP-SKFKDTYLVYILN	259

HMM_NAME Helicases conserved C-terminal domain

HMM	*EileeWLknlGlrvmYIHGdMpQeERdeIMddFNnGEynVLICtdVggr
-----	--

		++ + L+NLG++++ +HG+M+Q +R+ +++F++ +L++TDV++R	
Query	277	QRTALLLRNLGFTAIPHLHGQMSQSKRLGSLNKFKAARSILLATDVASR	325
HMM		GIDIPdVNVHVINYDMPWNPEqYIQRIGRTgRIG*	
		G+DIP V++V+N+D+P ++ +YI+R+GRT+R+G	
Query	326	GLDIPHDVVVNFDIPTHSKDYIHRVGRTARAG	358

DKF2phfbr2_71o20

group: brain derived

DKF2phfbr2_71o20 encodes a novel 232 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits
on genomic level encoded by AC006186 (3 exons)

Sequenced by GBF

Locus: /map="10q22.1"

Insert length: 1768 bp

Poly A stretch at pos. 1742, polyadenylation signal at pos. 1726

```
1 GGGGGCAGCA GGGCAAGGGG GAGGTGCGAG CGTGGACCTG GGACGGGTCT
51 GGGCGGGCTCT CGGTGGTTGG CACGGGTTGG CACACCCATT CAAGCGGCAG
101 GACGCACTTG TCTTAGCAGT TCTCGCTGAC CGCGCTAGCT GCGGCTTCTA
151 CGCTCCGGCA CTCTGAGTTC ATCAGCAAAC GCCCTGGCGT CTGTCTCTAC
201 CATGCCCTAGC CTTTGGGACC GCTTCTCGTC GTCGTCCACC TCCTCTTCGC
251 CCTCGTCTCT GCGCCGAAGT CCCACCCAGC ATCGGCCCGCC GCGCTCAGCC
301 TGGGGGTGGG CGACCCGGGA GGAGGGGTTT GACCGCTCCA CGAGCCTGGA
351 GAGCTCGGAC TCGGAGTCCC TGGACAGCAG CAACAGTGGC TTCGGGCCGG
401 AGGAAGACAC GGCTTACCTG GATGGGGTGT CGTTGCCCGA CTTGAGCTG
451 CTCAGTGACC CTGAGGATGA ACACTTGTGT GCCAACCTGA TGCAGCTGCT
501 GCAGGAGAGC CTGGCCAGG CGCGGCTGGG CTCTCGACGC CCTGCGCGCC
551 TGCTGATGCC TAGCCAGTTG GTAAGCCAGG TGGGCAAAGA ACTACTGCGC
601 CTGGCCTACA CCGAGCCGTG CGGCCTGCGG GGGCGCGTGC TGGACGTCTG
651 CGTGGAGCAG GGCAAGAGCT GCCACAGCGT GGGCCAGCTG GCACTCGACC
701 CCAGCCTGGT GCGCACCTTC CAGCTGACCC TCGTGCTGCG CCTGGACTCA
751 CGACTCTGGC CCAAGATCCA GGGGCTGTTT AGCTCCGCCA ACTCTCCCTT
801 CCTCCCTGGC TTCAGCCAGT CCCTGACGCT GAGCACTGGC TTCCGAGTCA
851 TCAAGAAGAA GCTGTACAGC TCGGAACAGC TGCCATTGA GGAGTGTGA
901 ACTTCAACCT GAGGGGGCGG ACAGTGCCTT CCAAGACAGA GACGACTGAA
951 CTTTGGGGT GGAGACTAGA GGCAGGAGCT GAGGGACTGA TTCCAGTGGT
1001 TGGAAAACCT AGGCAGCCAC CTAAAGTGGA GGTGGGGGAA TAGTGTTC
1051 CAGGAAGCTC ATTGAGTTGT GTGCGGTGGT CTGTGCATTG GGGACACATA
1101 CCCCTCAGTA CTGTAGCATG AAACAAAGGC TTAGGGGCCA ACAAGGCTTC
1151 CAGCTGGATG TGTGTGTAGC ATGTACCTTA TTATTTTGT TACTGACAGT
1201 TAACAGTGGT GTGACATCCA GAGAGCAGCT GGGCTGCTCC CGCCCCAGCC
1251 TGGCCAGGGG TGAAGGAAGA GGCACGTGCT CCTCAGAGCA GCCGGAGGGA
1301 AGGGGGAGGT CGGAGGTCGT GGAGGTGGTT TGTGTATCTT ACTGGTCTGA
1351 AGGGACCAAG TGTGTTTGT TTTGTTTGT TATCTGTTT TTCTGATCGG
1401 AGCATCACTA CTGACCTGTT GTAGGCAGCT ATCTTACAGA CGCATGAATG
1451 TAAGAGTAGG AAGGGGTGGG TGTCAGGGAT CACTTGGGAT CTTTGACACT
1501 TGAAAAATTA CACCTGGCAG CTGCGTTTAA GCCTTCCCCC ATCGTGTACT
1551 GCAGAGTTGA GCTGGCAGGG GAGGGGCTGA GAGGGTGGGG GCTGGAACCC
1601 CTTCCCGGGA GGAGTGCCAT CTGGGTCTTC CATCTAGAAC TGTTTACATG
1651 AAGATAAGAT ACTCACTGTT CATGAATACA CTTGATGTT AAGTATTAAG
1701 ACCTATGCAA TATTTTAC TTTTCTAATA AACATGTTT TTAACACAAA
1751 AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry AC006186 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 10 clone
CRI-JC2048 map 10q22.1; HTGS phase 1, 4 unordered pieces.

Score = 6512, P = 0.0e+00, identities = 1326/1345

3 exons

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 897 bp; peptide length: 232
Category: putative protein

```

1 MPSSLWDRFSS SSTSSSPSSL PRTPTPDRPP RSAWGSATRE EGFDRSTSLE
51 SSDCESLOSS NSGFGPEEDT AYLDGVSLPD FELLSDPEDE HLCANLMQLL
101 QESLAQARLG SRRPARLLMP SOLVSQVGKE LLRLAYSEPC GLRGALLDVC
151 VEQKKSCHSV GQLALDPSLV PTFQLTLVLR LDSRLWPKIQ GLFSSANSPP
201 LPGFSQSLTL STGFRVIKKK LYSSEQLPIE EC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_71o20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_71o20, frame 1

Report for DKFZphfbr2_71o20.1

```

[LENGTH]      232
[MW]           25354.60
[pI]           4.87
[PROSITE]      MYRISTYL          2
[PROSITE]      CK2_PHOSPHO_SITE    6
[PROSITE]      GLYCOSAMINOGLYCAN   1
[PROSITE]      PKC_PHOSPHO_SITE    1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      17.67 %

SEQ  MPSSLWDRFSSSTSSSPSSLPRTPTPDRPPRSAWGSATREEGFDRSTSLESSDCESLDSS
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  NSGFGPEEDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQESLAQARLGSRPARLLMP
SEG  xx.....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SOLVSQVGKELLRLAYSEPCGLRGALLDVCVEQKKSCHSVGQLALDPSLVPTFQLTLVLR
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LDSRLWPKIQGLFSSANSPPFLPGFSQSLTLSTGFRVIKKKLYSSEQLPIEEC
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_71o20.1

PS00002	62->66	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_71o20.1)

DKFZphfbr2_72b18

group: nucleic acid management

DKFZphfbr2_72b18 encodes a novel 715 amino acid protein with similarity to *E. coli* DNA-damage-inducible protein *dinP* and other proteins induced by DNA-damage.

The novel protein is similar to *dinP* of *E. coli*, *yqjH* of *B. subtilis*, *dinP* of *M. tuberculosis* and T19K24.15 of *A. thaliana*. The *dinB/P* pathway is a second SOS-pathway in *E. coli*. Therefore the new gene seems to be involved in DNA repair.

The new protein can find application in modulating DNA repair and mutagenesis.

similarity to DNA damage induced genes

complete cDNA, complete cds, potential start at Bp 49, EST hits
localisation primer site B is missing!

Sequenced by LMU

Locus: /map="416.0 cR from top of Chr18 linkage group"??

Insert length: 2475 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2431

```

1  GGGGGAGGAA  GGGGGCGGCG  ACGACGAGGA  AGACGCCGAG  GCCTGGGCCA
51  TGGAACTGGC  GGACGTGGGG  GCGGCAGCCA  GCTCGCAGGG  AGTTCATGAT
101  CAAGTGTTGC  CCACACCAAA  TGCTTCATCC  AGAGTCATAG  TACATGTGGA
151  TCTGGATTGC  TTTTATGCAC  AAGTAGAAAT  GATCTCAAAT  CCAGAGCTAA
201  AAGACAAACC  TTTAGGGGTT  CAACAGAAAT  ATTTGGTGGT  TACCTGCAAC
251  TATGAAGCTA  GGAACCTTGG  AGTTAAGAAA  CTTATGAATG  TCAGAGATGC
301  AAAAGAAAAA  TGTCCACAGT  TGGTATTAGT  TAATGGAGAA  GACCTGACCC
351  GCTACAGAGA  AATGTCTTAT  AAGGTTACAG  AATTACTGGA  AGAATTTAGT
401  CCAGTTGTTG  AGAGACTTGG  ATTTGATGAA  AATTTGTGG  ATCTAACAGA
451  AATGGTTGAG  AAGAGACTAC  AGCAGCTGCA  AAGTGATGAA  CTTTCTGCGG
501  TGAAGTGTGC  GGGTCATGTA  TACAATAATC  AGTCTATAAA  CCTGCTTGAC
551  GTCTTGACAC  TCAGACTACT  TGTGGATCT  CAGATTGCAG  CAGAGATGCG
601  GGAAGCCATG  TATAATCAGT  TGGGGCTCAC  TGGCTGTGCT  GGAGTGGCTT
651  CTAATAAACT  GTTGGCAAAA  TTAGTTTCTG  GTGCTTTTAA  ACCAAATCAA
701  CAAACAGTCT  TATTACCTGA  AAGTTGTCAA  CATCTTATTC  ATAGTTTGAA
751  TCACATAAAG  GAAATACCTG  GTATTGGCTA  TAAACTGCC  AAATGCTCTG
801  AAGCACTGGG  TATCAATAGT  GTGCGTGATC  TCCAAACCTT  TTCACCCAAA
851  ATTTTAGAAA  AAGAATTAGG  AATTTCAGTT  GCTCAGCGTA  TCCAAAAGCT
901  CAGTTTGGGA  GAGGATAACT  CCCCTGTGAT  ACTCTCAGGA  CCACCTCAGT
951  CCTTTAGTGA  AGAAGATTCA  TTTAAAAAAT  GTACATCTGA  AGTTGAAGCT
1001  AAAAATAAGA  TTGAAGAACT  ACTTGCTAGT  CTTTAAACA  GAGTATGCCA
1051  AGATGGAAGG  AAGCCTCATA  CAGTGAGATT  AATAATCCGT  CGGTATTCTT
1101  CTGAGAAGCA  CTATGGTCGT  GAGAGTCGTC  AGTGCCCTAT  TCCTTCACAT
1151  GTAATTCAGA  AATTAGGGAC  AGGAAATTAT  GATGTGATGA  CCCCATAGGT
1201  TGATATACTT  ATGAAACTTT  TTCGAAATAT  GGTGAATGTG  AAGATGCCAT
1251  TTCACCTTAC  CCTTCTAAGT  GTGTGCTTCT  GCAACCTTAA  AGCACTAAAT
1301  ACTGCTAAGA  AAGGGCTTAT  TGATTATTAT  TTAATGCCAT  CATTATCAAC
1351  TACTTCACGC  TCTGGCAAGC  ACAGTTTAA  AATGAAAGAC  ACTCATATGG
1401  AAGATTTTCC  CAAAGACAAA  GAAACAAACC  GGGATTTTCT  ACCAAGTGGA
1451  AGAATTGAAA  GTACAAGAAC  TAGGGAGTCT  CCACATAGATA  CCACAAATTT
1501  TTCTAAAGAA  AAAGACATTA  ATGAATTCCC  ACTCTGTTC  CTTCTGAAG
1551  GTGTTGACCA  AGAAGTCTCC  AAGCAGCTTC  CAGTAGATAT  TCAAGAAGAA
1601  ATCCTTTCTG  GAAATCTAG  GGAAAAATTT  CAAGGGAAAG  GAAGTGTGAG
1651  TTGTCCATTA  CATGCCCTCA  GAGGAGTATT  ATCTTTCTTT  TCTAAAAAAC
1701  AAATGCAAGA  TATTCCCATA  AATCCTAGAG  ATCATTTATC  CAGTAGCAAA
1751  CAGGTATCCT  CTGTATCTCC  TTGTGAACCG  GGAACATCAG  GCTTTAATAG
1801  CAGTAGTTCT  TCTTACATGT  CTAGCCAAAA  GGATTATTCA  TATTATTAG
1851  ATAATAGATT  AAAAGATGAA  CGAATAAGTC  AAGGACCTAA  AGAACCTCAA
1901  GGATTCCACT  TTACAAATTC  AAACCCTGCT  GTGTCTGCTT  TTCATTCTT
1951  TCCAAACTTG  CAGAGTGAGC  AACTTTTCTC  CAGAAACCAC  ACTACAGATA
2001  GCCATAAGCA  AACAGTAGCA  ACAGACTCTC  ATGAAGGACT  TACAGAAAAT
2051  AGAGAGCCAG  ATTCTGTTGA  TGAGAAAATT  ACTTTCCCTT  CTGACATTGA
2101  TCTTCAAGTT  TTCTATGAAC  TACCAGAAGC  AGTACAAAAG  GAACTGCTGG
2151  CAGAGTGGAA  GAGAACAGGA  TCAGATTTC  ACATTGGACA  TAAATAAGCA
2201  TATTACAGAA  AAAGGTCTGA  AAAGCAAGGG  AATACCATTA  TTTTCGGATT
2251  AGCGGTTTAT  TAAGCTCTTC  TATATTAAAC  ACTAATAGAT  ATTCAATAAC
2301  GGAGTAAACT  GTTCCAGATA  AAGCAAGAA  AGTTGCAAGA  AGTAAATCTT
2351  GGCACAAAGC  GTAAAAATAT  AACAGAAGAA  ATAATGTAAG  ATACTATCTT
2401  TTATGTCTAA  AGCCATTTTA  TATTACTTTT  CAATAAAAAG  AATATCATGG
2451  TCAAAAAAAA  AAAAAAATA  AAAAC

```

BLAST Results

Entry HS086339 from database EMBL:
human STS WI-11064.
Score = 1523, P = 3.0e-64, identities = 327/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2194 bp; peptide length: 715
Category: similarity to known protein

```
1 MELADVGAAG SSQGVHDQVL PTPNASSRVI VHVLDLCFYA QVEMISNPEL
51 KDKPLGVQOK YLVVTCNYEA RKLGVKKLMN VRDAKEKCPQ LVLVNGEDLT
101 RYREMSYKVT ELLEEFSPVV ERLGFDENFV DLTEMVEKRL QQLQSDLSA
151 VTVSGHVYNN QSINLLDVLH IRLVGSQIA AEMREAMYNQ LGLTGCAGVA
201 SNKLLAKLVS GVFKPNQQT VLLPESCQHLI HSLNHIKEIP GIGYKTAKCL
251 EALGINSVRD LQTFSPKILE KELGISVAQR IQKLSFGEDN SPVILSGPPPQ
301 SFSEEDSFKK CTSEVEAKNK IEELLASLLN RVCQDGRKPH TVRLIIRRY
351 SEKHYGRESR QCPIPSHVIQ KLGITGNYDVM TPMVDILMKL FRNMVNVKMP
401 FHLLTLLSVC CNLKALNTAK KGLIDYYLMP SLSTTSRSGK HSFKMKDTHM
451 EDFPKDKETN RDLFPGSRIE STRTRESPLD TTNFSKEKDI NEFFPLCSLPE
501 GVDQEVSKQL PVDIQEEILS GKSREKFQKG GSVSCPLHAS RGVLSFFSKK
551 QMQDIPINPR DHLSSSKQVS SVSPCEPGTS GFNSSSSSSYM SSQKDYSYYL
601 DNRLKDERIS QGPKEPQGFH FTNSNPAVSA FHSFPNLQSE QLFNRNHTTD
651 SHKQTVATDS HGLTENREP DSVDEKITFP SDIDPQVFYE LPEAVQKELL
701 AEWKRTGSDF HIGHK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72b18, frame 2

PIR:H64747 DNA-damage-inducible protein dinP - Escherichia coli, N =
2, Score = 212, P = 4.2e-27

PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis,
N = 2, Score = 230, P = 5.2e-26

>PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis
Length = 414

HSPs:

Score = 230 (34.5 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
Identities = 47/112 (41%), Positives = 73/112 (65%)

Query: 27 SRVIVHVDLCFYAQVEMISNPELKDKPLGV----QQKYLVVTCNYEARLGVKKLMNV 81
SR+I H+D++ FYA VEM +P L+ KP+ V ++K +VVTC+YEAR GVK M V
Sbjct: 5 SRIIFHIDMNSFYASVEMAYDPALRGKPVAVAGNVKERKGIIVTCSYEARARGVKTMPV 64

Query: 82 RDAKEKCPQLVLVNGEDLTRYREMSYKVTLEEFSPVVERLGFDEFVDLTE 134
AK CP+L+++ + RYR S + +L E++ +VE + DE ++D+T+
Sbjct: 65 WQAKRHCPQLVLP-PNFDYRNSSRAMFTILREYTDLVEPVSIDEGYMDMTD 116

Score = 137 (20.6 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
Identities = 43/148 (29%), Positives = 75/148 (50%)

Query: 178 QIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQT VLLPESCQHLIHSNLNHIK 237
+ A E++ + +L L G+A NK LAK+ S + KP T+L ++ L +
Sbjct: 125 ETAKIQSRQKELLPSIGIAPNKF LAKMASDMKKPLGITILRKQVPDILWPLP-VG 183

Query: 238 EIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSG 297
E+ G+G KTA+ L+ LGI+++ +L L++ LGI+ R++ + G ++PV
Sbjct: 184 EMHGVGKKTAELKGLGIHTIGELAADEHSLKRLGGIN-GPRLKNKANGIHAPV---- 238

Query: 298 PPQSFEEDSFKKCTSEVEAKNKIEELL 325
P+ E S ++ + EELL

Sbjct: 239 DPERIYEFKSVGNSSSTLSHDSSDEEELL 266

Pedant information for DKF2phfbr2_72b18, frame 2

Report for DKF2phfbr2_72b18.2

[LENGTH] 715
 [MW] 80300.63
 [pI] 6.37
 [HOMOL] TREMBL:SPBC16A3_11 gene: "SPBC16A3.11"; product: "hypothetical protein";
 S.pombe chromosome II cosmid c16A3. 5e-30
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YDR419w] 2e-15
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 genitalium, MG360] 3e-13
 [PIRKW] SOS mutagenesis 2e-11
 [PIRKW] DNA repair 2e-11
 [PIRKW] induced mutagenesis 2e-11
 [SUPFAM] umuC protein 3e-29
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] PROKAR_LIPOPROTEIN 1
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 21
 [PROSITE] ASN_GLYCOSYLATION 5
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 4.20 %

```

SEQ  MELADVGAASSQGVHDQVLPTPNASSRVIVHVDLDCFYAQVEMISNPELKDKPLGVQOK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  YLVVTCNYEARLGVKKLMNVDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPVV
SEG  .....
PRD  ceeeehhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc

SEQ  ERLGFDENFVDLTEMVEKRLQQLQSDLSAVTVSGHVVNNQSNINLLDLVHIRLLVGSQIA
SEG  .....
PRD  eecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  AEMREAMYNQLGLTGCVASNKLLAKLVSGVFKPNQQTVLLPESQHLIHSNLHIKEIP
SEG  .....
PRD  hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  GIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQ
SEG  .....
PRD  ccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFSEEDSFKKCTSEVEAKNIEELLASLLNRVCQDGRKPHTVRLIIRYSSEKHYGRESR
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QCPIPSHVIQKLGTGNYDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  KGLIDYLLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDLPSGRIESTRTRESPLD
SEG  .....
PRD  hhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TTNFSKEKDINEFPLCSLPEGVDQEVSKQLPVDIQEEILSGKSREKFQKGKSVSCPLHAS
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  RGVLSFFSKQMQDIPINPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSQKDYSYL
SEG  .....
PRD  hcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  DNRLKDERISQGPKEPQGFHTNSNPAVSFAHSPFNQSEQLFSRNHTTDSHKQTVATDS
SEG  .....
PRD  hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  HEGLTENREPDSDVEKITFPSDIDPQVFYELPEAVQKELLAEWKRTGSDFHIGHK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```


Prosites for DKFZphfbr2_72b18.2

PS00001	24->28	ASN_GLYCOSYLATION	PDOC00001
PS00001	160->164	ASN_GLYCOSYLATION	PDOC00001
PS00001	483->487	ASN_GLYCOSYLATION	PDOC00001
PS00001	583->587	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00004	309->313	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	347->351	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	106->109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	307->310	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	438->441	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	459->462	PKC_PHOSPHO_SITE	PDOC00005
PS00005	466->469	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	565->568	PKC_PHOSPHO_SITE	PDOC00005
PS00005	592->595	PKC_PHOSPHO_SITE	PDOC00005
PS00005	651->654	PKC_PHOSPHO_SITE	PDOC00005
PS00006	46->50	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	301->305	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	448->452	CK2_PHOSPHO_SITE	PDOC00006
PS00006	459->463	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	592->596	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	101->108	TYR_PHOSPHO_SITE	PDOC00007
PS00007	348->356	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	192->198	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	663->669	MYRISTYL	PDOC00008
PS00009	335->339	AMIDATION	PDOC00009
PS00013	186->197	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_72b18.2)

DKF2phfbr2_72d13

group: brain derived

DKF2phfbr2_72d13 encodes a novel 165 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

seems to be testis specific 9 of 10 EST hits are from testis libraries

Sequenced by LMU

Locus: unknown

Insert length: 723 bp

Poly A stretch at pos. 704, no polyadenylation signal found

```

1 AGGGGGGGTA TGGGGGAGGG GGAGACTCTG CAGGAGCCTA ATTCCCCACT
51 CTGAGCTCAC CCTTCTGTCT GCCCGGGGCC TACCCCTTCC CCTACTCTCA
101 CCCTTATAAT CCTTTTCAGC ACTAGGTCTT CCCGTCACCT CCACCTCTCT
151 CCATGACCCG GCTCTGCTTA CCCAGACCCG AAGCACGTGA GGATCCGATC
201 CCAGTTCCCTC CAAGGGGCCT GGGTGCTGGG GAGGGGTCAG GTAGTCCAGT
251 GCGTCCACCT GTATCCACCT GGGGCCCTAG CTGGGCCAG CTCCTGGACA
301 GTGTCCTATG GCTGGGGGCA CTAGGACTGA CAATCCAGGC AGTCTTTTCC
351 ACCACTGGCC CAGCCCTGCT GCTGCTTCTG GTCAGCTTCC TCACCTTTGA
401 CCGTCTCCAT AGGCCCGCAG GTCACACTCT GCCACAGCGC AAAGTTCTCA
451 CCAGGGGGCA GAGTCAGGGG GCCGGTGAAG GTCCTGGACA GCAGGAGGCT
501 CTACTCCTGC AAATGGGTAC AGTCTCAGGA CAACTTAGCC TCCAGGACGC
551 ACTGCTGCTG CTGCTCATGG GGCTGGGCCC GCTCCTGAGA GCCTGTGGCA
601 TGCCCTTGAC CCTGCTTGCC CTGGCTTTCT GCCTCCATCC TTGGGCCTGA
651 GAGCCCCCTC CCACAACCTA GTGTCCTTCA AATATACAAT GACCACCCTT
701 CTTCAAAAAA AAAAAAAAAA AAC

```

BLAST Results

Entry HS860F19 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 860F19
 Score = 2059, P = 1.1e-85, identities = 423/434
 2 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 153 bp to 647 bp; peptide length: 165
 Category: putative protein
 Classification: no clue

```

1 MTRLCLPRPE AREDPPIVPP RGLGAGEGSG SPVRPPVSTW GPSWAQLLDS
51 VLWLGLGLT IQAVFSTTGP ALLLLVLSFL TFDLLHRPAG HTLPQRKLLT
101 RGQSQGAGEG PGQQEALLLQ MGTVSGQLSL QDALLLLMG LGPLLRACGM
151 PLTLGLAFC LHPWA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_72d13, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72d13, frame 3

Report for DKFZphfbr2_72d13.3

[LENGTH] 165
[MW] 17393.73
[pI] 7.80
[BLOCKS] BL00068A Malate dehydrogenase proteins
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 29.70 %

SEQ MTRLCLPRPEAREDPIPVPPRGLGAGEGSGSPVRPPVSTWGPSWAQLLDSVLWLGAIGLT
SEG
PRD cchhhhhhhhhhhcccccc
MEM

SEQ IQAVFSTTGPAALLLLVSFLTFDLLHRPAGHTLPQRKLLTRGQSQAGEGPGQEQEALLQ
SEG
PRD eeeeeccccchhhhhhhhhhhhhccccccccccccccccccccccccccccchhhhhh
MEMMMMMMMMMMMMMMMMM.....

SEQ MGTVSGQLSLQDALLLLMGLGPLLRRACGMPLTLGLAFCLHPWA
SEG
PRD hccccccccchhhhhhhhhhhhhccccchhhhhccccchhhhhhhcccccc
MEMMMMMMMMMMMMMMMMM.....

(No Prosite data available for DKFZphfbr2_72d13.3)

(No Pfam data available for DKFZphfbr2_72d13.3)

DKFZphfbr2_72112

group: nucleic acid management

Summary DKFZphfbr2_72112 encodes a novel 344 amino acid protein with similarity to YDR126w and other *S. cerevisiae* proteins.

The novel protein contains a myc-type, helix-loop-helix dimerization domain signature. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, the protein could be a novel DNA-binding protein.

The new protein can application in modulating gene expression.

similarity to YDR126w ;
membrane regions: 2

similarity to YDR126w

complete cDNA complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1270 bp

Poly A stretch at pos. 1251, no polyadenylation signal found

```
1 GGGGGCGCCC GGGAGGCGCC GGAGCCAGC GGCTGGCGCC AGATCCAGGC
51 TCCTGGAAGA ACCATGTCCG GCAGCTACTG GTCATGCCAG GCACACACTG
101 CTGCCCAAGA GGAGCTGCTG TTTGAATTAT CTGTGAATGT TGGGAAGAGG
151 AATGCCAGAG CTGCCGGCTG AAAATTACCC AACCAAGAGA AATCTGCAGG
201 ATGGACTTTC TGGTCCTCTT CTGTCTCTAC CTGGCTTCGG TGCTGATGGG
251 TCTTGTTCTT ATCTGCGTCT GCTCGAAAAC CCATAGCTTG AAAGGCCTGG
301 CCAGGGGAGG AGCACAGATA TTTTCCTGTA TAATTCCAGA ATGTCTTCAG
351 AGAGCCGTGC ATGGATTGCT TCATTACCTT TTCCATACGA GAAACACAC
401 CTTCAATTGC CTGCACCTGG TCTTGCAAGG GATGGTTTAT ACTGAGTACA
451 CCTGGGAAGT ATTTGGCTAC TGTCAGGAGC TGGAGTTGTC CTTGCATTAC
501 CTTCTTCTGC CCTATCTGCT GCTAGGTGTA AACCTGTTTT TTTTACCCTT
551 GACTTGTGGA ACCAATCCTG GCATTATAAC AAAAGCAAAT GAATTATTAT
601 TTCTTCATGT TTATGAATTT GATGAAGTGA TGTTTCCAAA GAACGTGAGG
651 TGCTCTACTT GTGATTTAAG GAAACCAGCT CGATCCAAGC ACTGCAGTGT
701 GTGTAACCTG TGTGTGCACC GTTTCGACCA TCACTGTGTT TGGGTGAACA
751 ACTGCATCGG GGCCTGGAAC ATCAGGTAAT TCCTCATCTA CGTCTTGACC
801 TTGACGGCCT CGGCTGCCAC CGTCGCCATT GTGAGCACCA CTTTCTGGT
851 CCACTTGGTG GTGATGTCAG ATTATACCA GGAGACTTAC ATCGATGACC
901 TTGGACACCT CCATGTTATG GACACGGTCA TTCTTATTCA GTACCTGTTC
951 CTGACTTTTC CACGGATTGT CTTATGCTG GGCTTTGTCG TGGTCTTGAG
1001 CTTCTCTCTG GGTGGCTACC TGTGTCTGT CTTGTATCTG GCGGCCACCA
1051 ACCAGACTAC TAACGAGTGG TACAGAGGTG TCTGGGCTG GTGCCAGCGT
1101 TGTCCCCTTG TGGCCTGGCC TCCGTCAGCA GAGCCCCAAG TCCACCGGAA
1151 CATTCACTCC CATGGGCTTC GGAGCAACCT TCAAGAGATC TTTCTACCTG
1201 CCTTTCCATG TCATGAGAGG AAGAAACAAG AATGACAAGT GTATGACTGC
1251 CAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1232 bp; peptide length: 344
Category: similarity to unknown protein


```

1 MDFLVLFLFY LASVLMGLVL ICVCSKTHSL KGLARGGAQI FSCIPECLQ
51 RAVHGLLHYL FHTRNHTFIV LHLVLQGMVY TEYTWEVFGY CQELELSLHY
101 LLLPYLLLGVL NLFFFTLTG TNPGIITKAN ELLFLHVYEF DEVMFPKNVR
151 CSTCDLRKPA RSKHCSVCNW CVHRFDHHCW VVNNCIGAWN IRYFLIYVLT
201 LTASAATVAI VSTTFLVHLV VMSDLYQETY IDDLGHLHVM DTVILIQYLF
251 LTFPRIVFML GFVVVLSFLL GGYLLSVLYL AATNQTTNEW YRGVWAWCQR
301 CPLVAWPPSA EPQVHRNIHS HGLRSNLQEI FLPAFPCHER KKQE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72112, frame 3

TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1., N = 2, Score = 247, P = 1.4e-22

TREMBL:CED2021_3 gene: "D2021.2"; Caenorhabditis elegans cosmid
D2021., N = 1, Score = 209, P = 9e-17

TREMBL:CEC43H6_2 gene: "C43H6.7"; Caenorhabditis elegans cosmid
C43H6., N = 1, Score = 206, P = 5.2e-15

PIR:S52691 probable membrane protein YDR126w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 207, P = 8.4e-15

PIR:E71607 metal binding protein (DHHC domain) PFB0725c - malaria
parasite (Plasmodium falciparum), N = 1, Score = 182, P = 1.1e-13

>TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1.
Length = 356

HSPs:

Score = 247 (37.1 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 55/148 (37%), Positives = 85/148 (57%)

Query: 52 AVHGLLHYLFHTRNH--TFIVLHLVLQGM---VYTEYTWEVFGYCQELELSLHYLLLPY 105
A+ L +Y+ + N F+ L L+ G+ +Y + F + + L +LLPY
Sbjct: 64 AMRSLSNYVLYKNNPLVVFYLYLALITIGIASFFIYGSSLTQKFSIIDWISV-LTSVLLPY 122

Query: 106 LLLGVNLFFFTLTCGTNPGIITKANELLFLHVYEFD-EVMFPPKNVRCSTCDLRKPARSKH 164
++L+ + +NPG I N + +D ++ FP +CSTC KPARSKH
Sbjct: 123 ----ISLY---IAAKSNPGKIDLKNWNEASRRFPYDYKIFFPN--KCSTCKFEKPARSKH 173

Query: 165 CSVCNWCVHRFDHHCWVWNNCIGAWNIRYFLIYVL 199
C +CN CV +FDHHC+W+NNC+G N RYF +++L
Sbjct: 174 CRLCNICVEKFDHHCWVWNNCIGAWNIRYFLIYVL 208

Score = 43 (6.5 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 10/35 (28%), Positives = 17/35 (48%)

Query: 257 VFMLGFVV-VLSFLLGGYLLSVLYLAATNQTTNEW 290
VF++ + VL L GY ++Y T + +W
Sbjct: 254 VFLISLICSVLVLCLLGYEFFLVYAGYTTNESEKW 288

Pedant information for DKFZphfbr2_72112, frame 3

Report for DKFZphfbr2_72112.3

```

[LENGTH]      344
[MW]           39677.23
[pI]           7.26
[HOMOL]        TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein"; S.pombe
chromosome II cosmid c13G1. 3e-17
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YDR126w] 1e-16
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YDR264c] 8e-05
[FUNCAT]       10.05.99 other pheromone response activities (S. cerevisiae, YDR264c)
8e-05
[PIRKW]        transmembrane protein 4e-15
[SUPFAM]       ankyrin repeat homology 1e-10
[SUPFAM]       unassigned ankyrin repeat proteins 1e-10
[PROSITE]      MYRISTYL 4
[PROSITE]      CK2_PHOSPHO_SITE 3

```


Prosites for DKFZphfbr2_72112.3

(No Pfam data available for DKFZphfbr2_72112.3)

DKFZphfbr2_72m16

group: unknown

DKFZphfbr2_72m16 encodes a novel 287 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 1462 bp

Poly A stretch at pos. 1441, polyadenylation signal at pos. 1421

```
1 GGGGAGGACC GGAGGACCGA GGACAGAAAG ATTGGTGGAC AGGAGCAGCG
51 GCCCGTGGGG AGGGCGCTCG GCGGCGGCCT GCGGCCATGG CCACCGTGAT
101 GGCAGCGACG GCGGCGGAGC GGGCGGTGCT GGAGGAGGAG TTCCGCTGGC
151 TGCTGCACGA CGAGGTGCAC GCTGTGTTGA AGCAGCTGCA GGACATCCTC
201 AAGGAGGCCT CTCTGCGCTT CACTCTGCCG GGCTCCGGCA CTGAGGGGCC
251 CGCCAAGCAA GAGAACTTCA TCCTAGGCAG CTGTGGCACA GACCAGGTGA
301 AGGGTGTGCT GACTCTGCAG GGGGATGCCC TCAGCCAGGC GGATGTGAAC
351 CTGAAGATGC CCCGGAACAA CCAGCTGCTG CACTTCGCCT TCCGGGAGGA
401 CAAGCAGTGG AAGCTGCAGC AGATCCAGGA TGCCAGAAAC CATGTAGGCC
451 AAGCCATTTA CCTGCTTACC AGCCGGGACC AGAGCTACCA GTTCAAGACG
501 GCGCGTGAAG TCCTCAAGCT GATGGACGCA GTGATGCTGC AGCTGACCAG
551 AGCCCCGAAAC CGGCTCAGCA CCCCCGCCAC CCTCACCTC CCGGAGATCG
601 CCGCCAGCGG CCTCACGCGG ATGTTGCCCC CTGCCCTGCC GTCCGACCTG
651 CTGGTCAACG TCTACATCAA CCTCAACAAG CTCTGCCTCA CGGTGTACCA
701 GCTGCATGCC CTGCAGCCCA ACTCCACCAA GAACTTCCGC CCAGCTGGGG
751 GCGCGGTGCT GCATAGCCCT GGGGCCATGT TCGAGTGGGG CTCTCAGCGC
801 CTGGAGGTGA GCCACGTGCA CAAAGTGGAG TGGGTGATCC CCTGGCTCAA
851 CGACGCCCTG GTCTACTTCA CCGTCTCCCT GCAGCTCTGC CAGCAGCTTA
901 AGGACAAGAT TCCCGTGTTC TCCAGCTACT GGAGCTACAG ACCCTTCTGA
951 TCACAGCACC CAGGAGCTTG TCTCCAGGAA GCGGCCCCCG TCCCTACTC
1001 ATACCCACCA CAGAGCACCA GCCAGTGCCA ACGCCAGGCT GCTATTTATC
1051 TCCCTATCCC ACCCCCTACC CCACCTAACA CATTGCACT GCCGGGAATG
1101 GACACTGGAA GTGCCAGGAG GAAGGAAGGC TGGTTTGGTG GGGTAGTGGG
1151 GAGGTCAGGG AGGCGGGGCC AAGGGTGTCC CACATTCCCA ACACCGCCCT
1201 CTGATCACCA TGGGAATCTT TGGACTCAGG ACAGGGCCAG GCGCAGGGCT
1251 TCCCCTCCTC TCCCTTCGC TGTCCCTTCC CCTGGAGGG CATGGTGTGC
1301 GGGGGTGGCA CTGAGCTATG AGTCCCAGGG ATGCTGAGGA ACGCCACAGA
1351 CAGAGCCACC CTAGGAGTGA GTATAGTGCT GGTGACTGTG TTTATAGGCC
1401 CCAGTCCAGG GCTGTCTAAG AAATAAAGAT CATCAGACTC CAAAAAAGAA
1451 AAAAAAAGAA AC
```

BLAST Results

Entry HS604351 from database EMBL:
human STS WI-18474.
Score = 1178, P = 1.5e-48, identities = 250/268

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 87 bp to 947 bp; peptide length: 287
Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72ml6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72ml6, frame 3

Report for DKFZphfbr2_72m16.3

```
{LENGTH}      287
{MW}           32254.40
{pI}           8.30
{HOMOL}        TREMBL:AF025459 2 gene: "H14A12.3"; Caenorhabditis elegans cosmid H14A12. 3e-14
```

```
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 6
[PROSITE] PKC_PHOSPHO_SITE 5
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.27 %
```

[illegible]

```
SEQ      FTLGSCGTDVKGVLTLQGDALSQADVNLKMPRNNQLLHFAFREDKQWKLQQIQDARNHV
SEG      .....
PRD      hhhccccccceeeeeeeccccchhhhhhccccchhhhhhhhhcchhhhhhhhhhhchh
```

```
SEQ      SQAIYLLTSRDQSYQFKTGAEVCLKMDAVMLQLTRARNRLTPATLTLPEIAASGLTRMF  
SEG      .....  
PRD      hhhhhhccccceecchhhhhhhhhhhhhhhhhhhhhccccc
```

```

SEQ      APALPSDLLVNVIYNLNLKCLTVYQLHALQPNSTKNFRPAGGAVLHSPGAMFEWGSQRLE
SEG      .....
PRD      cccccccceeeehhhhhhhhhhhheeecccccccccccccccecccccccccccccee

```

```

SEQ      VSHVHKVECVI PWLNDALVYFTVSLQLCQQLKDKISVFSSYWSYRPF
SEG      .....
PRD      eeeeeeeeeeeccceeeeeeehhhhhhhhhhhhheeeeeeeccc

```

Prosites for DKFZphfbr2_72m16.3

PS000001	212->216	ASN_GLYCOSYLATION	PDOC000001
PS000005	42->45	PKC_PHOSPHO_SITE	PDOC000005
PS000005	128->131	PKC_PHOSPHO_SITE	PDOC000005
PS000005	213->216	PKC_PHOSPHO_SITE	PDOC000005
PS000005	236->239	PKC_PHOSPHO_SITE	PDOC000005
PS000005	283->286	PKC_PHOSPHO_SITE	PDOC000005
PS000006	8->12	CK2_PHOSPHO_SITE	PDOC000006
PS000006	50->54	CK2_PHOSPHO_SITE	PDOC000006
PS000006	83->87	CK2_PHOSPHO_SITE	PDOC000006
PS000006	128->132	CK2_PHOSPHO_SITE	PDOC000006
PS000006	138->142	CK2_PHOSPHO_SITE	PDOC000006
PS000006	167->171	CK2_PHOSPHO_SITE	PDOC000006
PS000008	64->70	MYRISTYL	PDOC000008

(No Pfam data available for DKF2phfbr2_72m16.3)

DKF2phfbr2_72n12

group: brain derived

DKF2phfbr2_72n12 encodes a novel 117 amino acid protein with similarity to a protein with conserved sequence in bacteria and eukariota.

The novel protein is very similar to human MM46, human and rat ganglioside expression factor-2 (GEF2), *C. elegans* 14.8 kD protein C32D5.9 and *Laccaria bicolor* symbiosis-related protein LBU93506_1. The function of this highly conserved proteins is not known.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to rat GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="12"

Insert length: 1880 bp

Poly A stretch at pos. 1859, polyadenylation signal at pos. 1830

```

1 GGGGGCCGGT ATTTCTCCAT CTGGCTCTCC TCTACCTCCA GGCAGGCTCA
51 CCCGAGATCC CCGCCCCGAA CCCCCCTGC AACTCGGCC CAGCGCTGTT
101 CCCCCCGGAG CGGACGTTTC TGCAGCTATT CTGAGCACAC CTTGACGTCG
151 GCTGAGGGAG CGGGACAGGG TCAGCGGCGA AGGAGGCAGG CCCCAGCGCGG
201 GGATCTCGGA AGCCTGCGG TGCATCATGA AGTTCCAGTA CAAGGAGGAC
251 CATCCCTTTG AGTATCGGAA AAAGGAAGGA GAAAAGATCC GGAAGAAATA
301 TCCGGACAGG GTCCCCGTGA TTGTAGAGAA GGCTCCAAAA GCCAGGGTGC
351 CTGATCTGGA CAAGAGGAAG TACCTAGTGC CCTCTGACCT TACTGTTGGC
401 CAGTTCTACT TCTTAATCCG GAAGAGAATC CACCTGAGAC CTGAGGACGC
451 CTTATTCTTC TTGTGCAACA ACACCATCCC TCCCACCAGT GCTACCATGG
501 GCCAACTGTA TGAGGACAAT CATGAGGAAG ACTATTTTCT GTATGTGGCC
551 TACAGTGATG AGAGTGTCTA TGGGAAATGA GTGGTTGGAA GCCCAGCAGA
601 TGGGAGCACC TGGACTTGGG GGTAGGGGAG GGGTGTGTGT GCGCGACATG
651 GGGAAAGAGG GTGGCTCCCA CCGCAAGGAG ACAGAAGGTG AAGACATCTA
701 GAAACATTAC ACCACACACA CCGTCATCAC ATTTTCACAT GCTCAATTGA
751 TATTTTTTGC TGCTTCCTCG GCCCAGGGAG AAAGCATGTC AGGACAGAGC
801 TGTGTGATTG GCTTTGATAG AGGAATGGGG ATGATGTAAG TTTACAGTAT
851 TCCTGGGGTT TAATTGTGTG GCAGTTTCAT AGATGGGTCA GGAGGTGGAC
901 AAGTTGGGGC CAGAGATGAT GGCAGTCCAG CAGCAACTCC CTGTGCTCCC
951 TTCTCTTTGG GCAGAGATTC TATTTTGTAC ATTTGCACAA GACAGGTAGG
1001 GAAAGGGGAC TTGTGGTAGT GGACCATACC TGGGGACCAA AAGAGACCCA
1051 CTGTAAATGA TGCATTGTGG CCCCTGATCT TCCCTGTCTC AACTTCTTTT
1101 TCCTCCATCC CGGTTGCAAT CTCACTCAGA CATCACAGTA CCACCCAGG
1151 GGTGGCAGTA GACAACAACC CAGAAATTTA GACAGGGATC TCTTACCTTT
1201 GGAATAATAG GGTAGGCAT GAAGGTGGTT GTGATTAAGA AGATGGTTTT
1251 GTTATTAAAT AGCATTAAAC TGAATTTGAC AAGAGTGTGT AGCATCCCTG
1301 TCTAACCTGC TCTTCTCTT TGGTGCCCTT TATCTACCC CTTCTTGGGA
1351 ATTTAATAAG TCTCAGGCAT TTCCAATTG AGACTAAAC CACTCTTAGC
1401 ATCTCCTCTA GTATTTTCCA TGTATCAGGA AAGAGGTGTC TTATGTAGGG
1451 AGGGGGCAAG TATGAAGTAA GGTAAATTATA TACTACTCTC ATTCAGGATT
1501 CTTGCTCCCA TGCTGCTGTC CCTTCAGGCT CACATGCACA GGAATGCTAC
1551 ATGATGGCCA GCTGCTTCCC TCCTTGGTTA TCATCCACTG CAGCTGCTAG
1601 TTAGAAAGGT TTGGAGGGAT GACTTTTAGT AAATCATGGG GATTTTATTG
1651 ATTTATTTTC ACTTTTGGGA TTTTGTGGGG TGGGAGTGGG GAGCAGGAAT
1701 TGCACTCAGA CATGACATTT CAATTCATCT CTGCTAATGA AAAGGGTTCT
1751 TTCTCTTGGG GGAAATGTGT GTGTCAGTTC TGTCAGCTGC AAGTTCTTGT
1801 ATAATGAAGT CAATGCCATC AGGCCAAGGA AATAAAATAA TTGCTTACCT
1851 TAAAAATCGA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HS418210 from database EMBL:
human STS SHGC-10496.

Score = 1916, P = 4.0e-80, identities = 394/400

Entry AC006514 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 68 unordered pieces.

Score = 610, P = 2.7e-16, identities = 128/134

4 exons

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 227 bp to 577 bp; peptide length: 117
 Category: strong similarity to known protein

1 MKFQYKEDHP FEYRKKEGEG IRKKYPDRVP VIVEKAPKAR VPDLDKRKYL
 51 VPSDLTVGQF YFLIRKRIHL RPEDALFFV NNTIPPTSAT MGQLYEDNHE
 101 EDYFLYVAYS DESVYGK

BLASTP hits

Entry YQD9 CAEL from database SWISSPROT:
 HYPOTHETICAL 14.8 KD PROTEIN C32D5.9 IN CHROMOSOME II.
 Score = 496, P = 1.8e-47, identities = 91/116, positives = 105/116

Entry SYRP LACBI from database SWISSPROT:
 SYMBIOSIS-RELATED PROTEIN.
 Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry LBU93506.1 from database TREMBL:
 product: "symbiosis-related protein"; Laccaria bicolor
 symbiosis-related protein mRNA, partial cds.
 Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry GEF2_RAT from database SWISSPROT:
 GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2).
 Score = 373, P = 2.0e-34, identities = 71/116, positives = 88/116

Alert BLASTP hits for DKFZphfbr2_72n12, frame 2

TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
 cds., N = 1, Score = 549, P = 4.7e-53

SWISSPROT:GEF2_HUMAN GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)., N = 1,
 Score = 373, P = 2.1e-34

>TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
 cds.
 Length = 117

HSPs:

Score = 549 (82.4 bits), Expect = 4.7e-53, P = 4.7e-53
 Identities = 101/116 (87%), Positives = 110/116 (94%)

Query: 1 MKFQYKEDHPFEYRKKEGEGIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPDLTVGQF 60
 MKF YKE+HPFE R+ EGKIRKKYPDRVPVIVEKAPKAR+ DLDK+KYLVPDLTVGQF
 Sbjct: 1 MKFVYKEEHPFEKRRSEGEKIRKKYPDRVPVIVEKAPKARIGDLDKKYLVPDLTVGQF 60

Query: 61 YFLIRKRIHLRPEDALFFVNNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYG 116
 YFLIRKRIHLR EDALFFVNN IPPTSATMGQLY+++HEED+FLY+AYSDESVYG
 Sbjct: 61 YFLIRKRIHLRAEDALFFVNNVIPPTSATMGQLYQEHEEDFFLYIAYSDESVYG 116

Pedant information for DKFZphfbr2_72n12, frame 2

Report for DKFZphfbr2_72n12.2

[LENGTH] 117
 [MW] 14044.07
 [pI] 8.67
 [HOMOL] TREMBL:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete cds. 1e-56

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YBL078c] 4e-36
[SUPFAM] hypothetical protein YBL078c 8e-35
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDRKKYLVPDLTVGQF
PRD cccccccccchhhhhhhhhhhhhccccceeeccccccccccccccccccccccccchhh

SEQ YFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESIVYGK
PRD hhhhhhhhhccccceeeccccccccchhhhhhhhhccccceeecccccccccccc

Prosite for DKFZphfbr2_72n12.2

PS00001 81->85 ASN_GLYCOSYLATION PDOC00001

(No Pfam data available for DKFZphfbr2_72n12.2)

DKFZphfbr2_78c24

group: signal transduction

DKFZphfbr2_78c24 encodes a novel 563 amino acid protein with strong similarity to guanylate-binding proteins (GBPs).

GBPs were originally described as proteins that are strongly induced by interferons and are capable of binding to agarose-immobilized guanine nucleotides. hGBP1, the first of two members of this protein family in humans, represents a novel type of GTPase. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and a RGD cell attachment site. It seems to be a new member of the GBP-family and shows a splicing pattern not described previously.

The new protein can find application in modulating/blocking the response of cells to interferons.

strong similarity to guanine nucleotide-binding protein 1/2
but different "splice variant" aa 211-245 of GBP1/2 missing

Sequenced by MediGenomix

Locus: unknown

Insert length: 2952 bp

Poly A stretch at pos. 2927, polyadenylation signal at pos. 2914

```
1 CAGTTTCATT AGGCTCTGAA GCCATTACAA AGGTTGCTTA ACTTCTAATT
51 ATTTGATCAC TGAGGAAAAT CCAGAAAGCT ACACAACACT GAAGGGGTGA
101 AATAAAAGTC CAGCGATCCA GCGAAAGAAA AGAGAAGTGA CAGAAACAAC
151 TTTACCTGGA CTGAAGATAA AAGCACAGAC AAGAGAACAA TGCCTTGGAC
201 ATGGCTCCAG AGATCCACAT GACAGGCCCA ATGTGCCTCA TTGAGAACAC
251 TAATGGGGAA CTGGTGGCGA ATCCAGAAGC TCTGAAAATC CTGTCTGCCA
301 TTACACAGCC TGTGGTGGTG GTGGCAATTG TGGGCCTCTA CCGCACAGGA
351 AAATCCTACC TGATGAACAA GCTAGCTGGG AAGAATAAGG GCTTCTCTCT
401 GGGCTCCACA GTGAAATCTC ACACCAAAGG AATCTGGATG TGGTGTGTGC
451 CTCACCCCAA AAAGCCAGAA CACACCTTAG TCCTGCTTGA CACTGAGGGC
501 CTGGGAGATG TAAAGAAGGG TGACAACCAG AATGACTCCT GGATCTTCAC
551 CCTGGCCGTC CTCCTGAGCA GCACTCTCGT GTACAATAGC ATGGGAACCA
601 TCAACCAAGC GGCTATGGAC CAACTGTACT ATGTGACAGA GCTGACACAT
651 CGAATCCGAT CAAAATCCTC ACCTGATGAG AATGAGAATG AGGATTTCAGC
701 TGACTTTGTG AGCTTCTTCC CAGATTTTGT GTGGACACTG AGAGATTCTCT
751 CCCTGGACTT GGAAGCAGAT GGACAACCCC TCACACCAGA TGAGTACCTG
801 GAGTATTCCC TGAAGCTAAC GCAAGGTAAC AGGAAGCTTG CCCAGCTTGA
851 GAAACTACAA GATGAAGAGC TGGACCCTGA ATTTGTGCAA CAAGTAGCAG
901 ACTTCTGTTC CTACATCTTT AGCAATTCCA AAATAAAAC TCTTTCAGGA
951 GGCATCAAGG TCAATGGGCC TTGTCTAGAG AGCCTAGTGC TGACCTATAT
1001 CAATGCTATC AGCAGAGGGG ATCTGCCCTG CATGGAGAAC GCAGTCCCTGG
1051 CCTTGGCCCA GATAGAGAAC TCAGCCGAGC TGCAAAAGGC TATTGCCCAC
1101 TATGACCAGC AGATGGGCCA GAAGGTGCAG CTGCCCGCAG AAACCCCTCA
1151 GGAGCTGCTG GACCTGCACA GGGTTAGTGA GAGGGAGGCC ACTGAAGTCT
1201 ATATGAAGAA CTCTTCAAG GATGTGGACC ATCTGTTTCA AAAGAAATTA
1251 GCGGCCCAGC TAGACAAAAA GCGGGATGAC TTTTGTAAAC AGAATCAAGA
1301 AGCATCATCA GATCGTTGCT CAGCTTTACT TCAGGTCATT TTCAGTCTC
1351 TAGAAGAAGA AGTGAAGGCG GGAATTTATT CGAAACCAGG GGGCTATTGT
1401 CTCTTTATTC AGAAGCTACA AGACCTGGAG AAAAAGTACT ATGAGGAACC
1451 AAGGAAGGGG ATACAGGCTG AAGAGATTCT GCAGACATAC TTGAAATCCA
1501 AGGAGTCTGT GACCGATGCA ATTCTACAGA CAGACCAGAT TCTCACAGAA
1551 AAGGAAAAGG AGATTGAAGT GGAATGTGTA AAAGCTGAAT CTGCACAGGC
1601 TTCAGCAAAA ATGGTGGAGG AAATGCAAAAT AAAGTATCAG CAGATGATGG
1651 AAGAGAAAAG GAAGAGTTAT CAAGAACATG TGAACAATT GACTGAGAAG
1701 ATGGAGAGGG AGAGGGCCCA GTTGCTGGAA GAGCAAGAGA AGACCTCAC
1751 TAGTAACTT CAGGAACAGG CCCGAGTACT AAAGGAGAGA TGCCAAGGTG
1801 AAAGTACCCA ACTTCAAAAT GAGATACAAA AGCTACAGAA GACCCTGAAA
1851 AAAAAAACCA AGAGATATAT GTCGCATAAG CTAAAGATCT AAACAACAGA
1901 GCTTTTCTGT CATCCTAACC CAAGGCATAA CTGAAACAAT TTTAGAATTT
1951 GGAACAAGTG TCACTATATT TGATAATAAT TAGATCTTGC ATCATAACAC
2001 TAAAAGTTTA CAAGAACATG CAGTTCAATG ATCAAAATCA TGTTTTTCC
2051 TAAAAGAT TGTAATTGT GCAACAAAGA TGCAATTTACC TCTGTACCAA
2101 CAGAGGAGGG ATCATGAGTT GCCACCACTC AGAAGTTTAT TCTTCCAGAC
2151 GACCAAGTGA TACTGAGGAA AGTCTTAGGT AAAAATCTTG GGACATATTT
2201 GGGCACTGGT TTGGCCAAGT GTACAATAGG TCCCAATATC AGAAACAACC
2251 ATCCTAGCTT CCTAGGGAAG ACAGTGTACA GTTCTCCATT ATATCAAGGC
2301 TACAAGGTCT ATGAGCAATA ATGTGATTTT TGGACATTGC CCATGGATAA
2351 TTCTCACTGA TGGATCTCAA GCTAAAGCAA ACCATCTTAT ACAGAGATCT
2401 AGAATCTTAT ATTTCCATA GGAAGGTAAA GAAATCATTA GCAAGAGTAG
2451 GAATTGAATC ATAAACAAAT TGGCTAATGA AGAATCTTT TCTTCTTGT
2501 TCAATTCATC TAGATTATAA CCTTAATGTG ACACCTGAGA CCTTTAGACA
```



```

2551 GTTGACCCCTG AATTAAATAG TCACATGGTA ACAATTATGC ACTGTGTAAT
2601 TTTAGTAATG TATAACATGC AATGATGCAC TTTAACTGAA GATAGAGACT
2651 ATGTTAGAAA ATTGAACATA TTTAATTATT TGATTGTTTT AATCCTAAAG
2701 CATAAGTTAG TCTTTTCCTG ATTCTTAAAG GTCATACTTG AAATCCTGCC
2751 AATTTTCCCC AAAGGGAATA TGGAAATTTT TTTGACTTTC TTTTGAGCAA
2801 TAAAATAAAT GTCTTGCCAT TACTTAGTAT ATGTAGACTT CATCCCAATT
2851 GTCAACATC CTAGGTAAGT GGTGACATT TCTTACAGCA ATTACAGATT
2901 ATTTTGAAC TAGAAATAAA CTAACTAGA AACAAAAAAA AAAAAAAA
2951 AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1889 bp; peptide length: 563
 Category: strong similarity to known protein
 Classification: Cell signaling/communication
 Prosite motifs: RGD (272-275)
 ATP_GTP_A (45-53)

```

1 MAPEIHMTGP MCLIENTNGE LVANPEALKI LSAITQPVVV VAIVGLYRTG
51 KSYLMNKLKAG KNGFSLGST VKSHTKGIWM WCVPHPKKPE HTLVLLDTEG
101 LGDVKKGDNQ NDSWIFTLAV LLSSTLVYNS MGTINQQAMD QLYYVTELT
151 RIRSKSSPDE NENEDSADFV SFFPDFVWTL RDFSLEAD GQPLTPDEYL
201 EYSLKLTQGN RKLAQLEKLQ DEELDPEFVQ QVADFCSYIF SNSKTKTSLG
251 GIKVNGPCLE SLVLTYNINAI SRGDLPCMEN AVLALAQIEN SAAVQKAIH
301 YDQMGQKQV LPAETLQELL DLHRVSEREA TEVYMKNSFK DVDHLEFQKKL
351 AAQDKKRDD FCKQNEASS DRCSALLQVI FSPLEEEVKA GIYSKPGGYC
401 LFIQKLQDLE KKYEEPRKG IQAEELQTY LKSKEVTDI ILQTDQILTE
451 KEKEIEVECV KAESAQASAK MVEEMQIKYQ QMEEKEKSY QEHVKQLTEK
501 MERERAQLE EQEKLTSKL QEQARVLKER CQGESTQLQN EIQLQKTLK
551 KKTKRYMSHK LKI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78c24, frame 3

PIR:A41268 guanine nucleotide-binding protein 1 - human, N = 2, Score = 1306, P = 4.9e-238

PIR:A46459 macrophage-activation gene-1 protein mag-1 - mouse, N = 2, Score = 942, P = 8.9e-184

PIR:S70524 guanine nucleotide-binding protein 2 - human, N = 2, Score = 1131, P = 4.1e-210

TREMBL:AF077007_1 gene: "Gbp2"; product: "interferon-induced guanylate binding protein GBP-2"; Mus musculus interferon-induced guanylate binding protein GBP-2 (Gbp2) mRNA, complete cds., N = 2, Score = 904, P = 1.2e-179

>PIR:A41268 guanine nucleotide-binding protein 1 - human
 Length = 592

HSPs:

Score = 1306 (195.9 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
 Identities = 264/332 (79%), Positives = 288/332 (86%)

Query: 211 RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTSLGGIKVNGPCLESVLVTYINAI 270
 RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTSLGGI+VNGP LESVLVTY+NAI
 Sbjct: 245 RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTSLGGIQVNGPRLESVLVTYVNAI 304

Query: 271 SRGDLPCMENAVLALAQIENSAAVQKAIHAHYDQMGQKVQLPAETLQELLDLHRVSEREA 330
S GOLPCMENAVLALAQIENSAAVQKAIHAHY+QMGQKVQLP E+LQELLDLHR SEREA
Sbjct: 305 SSGDLPCMENAVLALAQIENSAAVQKAIHAHYEQMGQKVQLPTESLQELLDLHRDSEREA 364

Query: 331 TEVYMKNFSKDVDFHFLQKLLAAQLDKRRDDFCQKQNEASSDRCSALLQVIFSPLEEEVKA 390
EV+++SFKDVDFHFLQK+LAAQL+KKRRDDFCQKQNEASSDRCS LLQVIFSPLEEEVKA
Sbjct: 365 IEVFIRSSFKDVDFHFLQKELAAQLEKKRRDDFCQKQNEASSDRCSGLLQVIFSPLEEEVKA 424

Query: 391 GIYSKPGGYCLFIQKLQDLEKKYEEPRKGIQAEELQTYLKSKESTDAILOTDQILTX 450
GIYSKPGGY LF+QKLQDL+KKYEEPRKGIQAEELQTYLKSKESTDAILOTDQ LT
Sbjct: 425 GIYSKPGGYRLFVQKLQDLKKYEEPRKGIQAEELQTYLKSKESTDAILOTDQTLTE 484

Query: 451 XXXXXXXXXXXXXSAQASAKMVEEMQIKYQMMEEKEKSYQEHVKQLTEKXXXXXXXXXX 510
SAQASAKM++EMQ K +QME+KE+SYQEH+KQLTEKM
Sbjct: 485 KEKEIEVERVKAESAQASAKMLQEMQRKNEQMEQKERSYQEHKQLTEKMENDRVQLLK 544

Query: 511 XXXKTLTSLKQEQARVLKERCQGESTQLQNEI 542
+TL KLQEQ ++LKE Q ES ++NEI
Sbjct: 545 EQERTLALKLQEQQLLKEGFQKESRIMKNEI 576

Score = 1012 (151.8 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
Identities = 194/211 (91%), Positives = 200/211 (94%)

Query: 1 MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLKAG 60
MA EIHMTGPMCLIENTNG L+ANPEALKILSAITQP+VVVVAIVGLYRTGKSYLMNKLKAG
Sbjct: 1 MASEIHMTGPMCLIENTNGRLMANPEALKILSAITQPMVVVAIVGLYRTGKSYLMNKLKAG 60

Query: 61 KKNKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNDQNDSWIFTLAV 120
K KGFSLGSTV+SHTKGIWMWCVPHPKK H LVLLDTEGLGDV+KGDNDQNDSWIF LAV
Sbjct: 61 KKKGFSLGSTVQSHTKGIWMWCVPHPKKPGHILVLLDTEGLGDVEKGDNDQNDSWIFALAV 120

Query: 121 LLSSTLVYNSMGITINQQAMDQLYYVTELTHRIRSKSSPDENENE--DSADFVSFFPDEFVW 178
LLSST VYNS+GTINQQAMDQLYYVTELTHRIRSKSSPDENENE DSADFVSFFPDEFVW
Sbjct: 121 LLSSTFVYNSIGITINQQAMDQLYYVTELTHRIRSKSSPDENENEVEDSADFVSFFPDEFVW 180

Query: 179 TLRDFSLDLEADGQPLTPDEYLEYSLKLTQ 209
TLRDFSLDLEADGQPLTPDEYL YSLKL +G
Sbjct: 181 TLRDFSLDLEADGQPLTPDEYLTYSKLKKG 211

Pedant information for DKFZphfbr2_78c24, frame 3

Report for DKFZphfbr2_78c24.3

[LENGTH] 563
[MW] 64127.72
[PI] 5.45
[HOMOL] PIR:A41268 guanine nucleotide-binding protein 1 - human 0.0
[SUPFAM] guanine nucleotide-binding protein 1 0.0
[PROSITE] ATP_GTP_A 1
[PROSITE] RGD 1
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 6.75 %
[KW] COILED_COIL 10.48 %

SEQ MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLKAG
SEG
PRD cccccccccccccccccchhhhhhhhhhhhhccccccccccccccchhhhhhhhh
COILS
MEMMMMMMMMMMMMMMMMM

SEQ KKNKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNDQNDSWIFTLAV
SEG
PRD cchhhhhhhhh
COILS
MEM

SEQ LLSSTLVYNSMGITINQQAMDQLYYVTELTHRIRSKSSPDENENEDSADFVSFFPDEFVWTL
SEG
PRD hhhhheccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccc
COILS
MEM

SEQ RDFSLEADGQPLTPDEYLEYSLKLTQGNRKLQLEKLQDEELDPEFVQQVADFCSYIF
SEG
PRD hhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhc
COILS


```
MEM .....
SEQ  SNSKTKTLSGGIKVNPGCLESVLTYINAI SRGDLPCMENAVLALA QIENSAAVQKATIAH
SEG  .....
PRD  cccceeeccccccccccccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  YDQMGQKVQLPAETLQELLDLHRVSE REATEVYMKN SFKDVDHLFQKKLAAQLDKKRDD
SEG  .....
PRD  hhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  FCKQNEASSDRCSALLQVIFSPLEEEVKAGIYSKPGGYCLFIQKLQDLEKKYYEEPRKG
SEG  .....
PRD  hhhhhhhchhhhhhhhhhhhhhhhhhhhhhhccccccccceehhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  IQAEIILQTYLKSKE SVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIKYQ
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  QMMEEEKSYQEHVKQLTEKMERERAQLLEEQEKLTLSKLQEQARVLKERCGESTQLQN
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCC
MEM  .....

SEQ  EIQLQKTLKKKTKRYMSHKLKI
SEG  ..xxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCC.....
MEM  .....
```

Prosites for DKF2phfbr2_78c24.3

PS00016	272->275	RGD	PDOC00016
PS00017	45->53	ATP_GTP_A	PDOC00017

(No Pfam data available for DKF2phfbr2_78c24.3)

DKFZphfbr2_78d13

group: brain derived

DKFZphfbr2_78d13 encodes a novel 259 amino acid protein with similarity to *C. elegans* putative protein from cosmid K08B12.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *C.elegans* K08B12.3

Sequenced by MediGenomix

Locus: /map="338.4 cR from top of Chr18 linkage group"

Insert length: 2195 bp

Poly A stretch at pos. 2175, polyadenylation signal at pos. 2156

```
1 CGTCCGTCGG GCAGCAGCGG GGCTGTCTAT CCCGGCTGAG GACCCGCGGC
51 CAGTGCCGGT GGCTGGCTTT GCCATTAGCG GGGGCCTTTC CTGAGGACGG
101 CGTACGGAGT GTGGGGAATG AAGGATGGCA GCATGCCGTG CATTAAAAGC
151 TGTTTTGGTA GATCTCAGTG GCACACTTCA CATTGAAGAT GCAGCTGTGC
201 CAGGCGCACA GGAAGCTCTT AAAAGGTTAC GTGGTGCTTC TGTAATCATT
251 AGGTTTGTGA CCAATACAAC CAAAGAGAGC AAGCAAGACC TGTAGAAAG
301 GTTGAGAAAA TTGGAATTG ATATCTCTGA AGATGAAATA TTCACATCTC
351 TGA CTGCAGC CAGAAGTTTA CTAGAGCGGA AACAACTCAG ACCCATGCTG
401 CTAGTTGATG ATCGGGCACT ACCTGATTTC AAAGGAATAC AAACAAGTGA
451 TCCTAATGCT GTGGTCATGG GATTGGCACC AGAACATTTT CATTATCAAA
501 TTCTGAATCA AGCATTCCGG TTACTCCTGG ATGGAGCACC TCTGATAGCA
551 ATCCACAAAG CCAGGTATTA CAAGAGGAAA GATGGCTTAG CCCTGGGGCC
601 TGGACCATTT GTGACTGCTT TAGAGTATGC CACAGATACC AAAGCCACAG
651 TCGTGGGGAA ACCAGAGAAG ACGTTCTTTT TGGAGCATT GCGGGGCACT
701 GGCTGTGAAC CTGAGGAGGC TGTCATGATA GGAGATGATT GCAGGGATGA
751 TGTGGTGGG GCTCAAGATG TCGGCATGCT GGGCATCTTA GTAAAGACTG
801 GGAATATCG AGCATCAGAT GAAGAAAAAA TTAATCCACC TCCTTACTTA
851 ACTTGTGAGA GTTTCCTTCA TGCTGTGGAC CACATTCTGC AGCACCTATT
901 GTGAAGCAAT GTGTGCATCT GAAGCAACTT GAAATGCAGC TTCTTATTGT
951 CTGGAATGAA TCCCTTACCA ACTCAGTGCC AGCATCGGTA GACACCAGTC
1001 AGTGCTGATC GCTTTTAAAC CCTCTTTTGT TGTGCATTAA TTAGAAAGAA
1051 AGGTATTGAA TTGCGGCTAG CCAGTAAGCC TTGCTAATCT CTTTTATTTT
1101 GTAAC TGAAG ATGAGACCCA AAGAAAGGGA AAGCTGAGAT TTTGTGCCAT
1151 TCCTTTTAAA ATATTATCA GGTAGGTGG GGCTGTGGGG GAAAAGCTAC
1201 TACAGGGAAG AGTGTCTCTT GCTGTCTCTT CACTGGAAAA CAGGGAGGGG
1251 GGATTTTACA CTGTGAAGAA AGTTGAATGG TGGTTTTTAA ATTATAAAGT
1301 AATGTATTAA AAGGTGCATT AGGCTGTAGT TCTAATATTG AGTCAACTG
1351 TGAAATCCAT CAGATGTGCC AAATGGAGAA GACAGAAAGC AACAAAGTGA
1401 ATTGTTCTTT AGCCCAAGTG GTACAGTGAA TTTGCTTTAA CAGATGTTGA
1451 AAAC TAAATT TTCTACTGTA TTCCAGCAC GGGTGACTTC TTTTCTCTT
1501 CATTAGCCAG AGATGACTAA TTTAAATTTA GAACCAGATT TTAATTTAAA
1551 TTAATATTTC CATTAATAAC CTACTCATTG CAGATACCTA TTACTACTGTG
1601 TAACAGTTGT TTTGGAATTT TTATGTAAAA TTAACACTAT CAGTATTTTA
1651 CAGATGTTTT AATTAGACAT TGTTATTAA AGGAACAGTG CAGAACTAG
1701 AATCAAGCCT TATAATATCT TATAGCCAT GCATTTTGA AGTTAGTGTC
1751 CACTAGGGTC CTATTAAC TGACATTGCA AGATTTTATT ATTTTGCCT
1801 CTGACACTAT GGGAAAAATT TTTAGAAGC TATTGGGACA GATTCAAGCT
1851 TTTATGCACT TGGTTACTAC AGCTGTAAAA TGAAATCTCG TCTGTAGCA
1901 TGGATTATTC TTCTCATGTT AAACCCACCA AAATAAGGG GACTAAATAG
1951 GTAATGATTT TCCTAGTGCA TTTGCATACT GTGATAATCC TGGGCCTTGC
2001 AATAGTTCTA CAGGGCTCTT GGGCATTGAA TTATTAGGAT GTAATTGTAC
2051 ATCATTGTAG TGTTACCTT ATTGAAGCTC ACTCTGATGT TAATGAGCTT
2101 CCGGTTTGA TGCTTGTTTA GAGATCAGCA GTCTTGGATG GGAGGGAACA
2151 AAGCTAAATA AATGTTAGTT TGGTGAAAAA AAAAAAAA AAAAA
```

BLAST Results

Entry HS599355 from database EMBL:

human STS WI-13484.

Score = 1262, P = 3.6e-52, identities = 274/289

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 901 bp; peptide length: 259
Category: similarity to unknown protein
Classification: no clue

```

1 MAACRALKAV LVDLSGTLHI EDAAVPGAQE ALKRLRGASV IIRFVNTTTK
51 ESKQDLLERL RKLEFDISED EIFTSLTAAR SLLERKQVRP MLLVDDRALP
101 DFKGIQTSDP NAVVMGLAPE HFHYQILNOA FRLLLDGAPL IAIHKARYYK
151 RKDGLALGPG PFVTALEYAT DTKATVVGKP EKTFFLEALR GTGCEPEEAV
201 MIGDDCRDDV GGAQDVGMGL ILVKTGKYRA SDEEKINPPP YLTCESEFPA
251 VDHILQHLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78d13, frame 2

TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid
K08B12., N = 1, Score = 609, P = 2.2e-59

TREMBL:CEC13C4_5 gene: "C13C4.4"; Caenorhabditis elegans cosmid C13C4,
N = 1, Score = 408, P = 4.4e-38

>TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid
K08B12.
Length = 257

HSPs:

Score = 609 (91.4 bits), Expect = 2.2e-59, P = 2.2e-59
Identities = 132/251 (52%), Positives = 172/251 (68%)

```

Query:      7 LKAVLVDSLGLHIEDAAVPGAQEALKRLRGASVIIRFVNTTTKESQDLLERLRKLEFD 66
             + +VL+DLSGT+HIE+ A+PGAQ AL+ LR + + +FVTNTTTKESK+ L +RL F
Sbjct:      4 ISSVLIDLSGTIHIEEFAIPGAQTALELLRQHAKV-KFVTNTTTKESKRLLHQRLINCGFK 62

Query:     67 ISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTSDPNAVVMGLAPEHFHYQI 126
             + ++EIFTSLTAAR L+ + Q RP +VDDRA+ DF+GI T DPNVAV+GLAPE F+
Sbjct:     63 VEKEEIFTSLTAARDLIVKNQYRPFVIVDDRAMEDFEGISTDDPNVAVVIGLAPEKFNDTT 122

Query:    127 LNQAERLLLDG-APLIAHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKPEKTF 185
             L AERL+ + A LIAI+K RY++ GL LGPG +V LEY+ +AT+VGKP K FF
Sbjct:    123 LTHAFRLIKEKASLIAINKGRYHQTNAGLCLGPGTYVAGLEYSAGVEATIVGKPNKLFF 182

Query:    186 LEALRGTG--CEPEEAVMIGDDCRDDVGAQDVGMGLILVKTGKYRASDEEKINPPPYLT 243
             AL+ + AVMIGDD DD GA +GM ILVKTGK+R DE K+
Sbjct:    183 ESALQSLNENVDFSSAVMIGDDVNDALGAIGMRAILVKTGKFRDGDDELKVKN----V 238

Query:    244 CESFPHAVDHILQH 257
             SF AV+ I+++
Sbjct:    239 ANSFVDAVNMIEN 252

```

Pedant information for DKFZphfbr2_78d13, frame 2

Report for DKFZphfbr2_78d13.2

```

[LENGTH]      259
[MW]           28536.04
[pI]           5.84
[HOMOL]        TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12. 3e-
62
[FUNCAT]       r general function prediction [M. jannaschii, MJ1437] 3e-05
[SUPFAM]       nagD protein 4e-18
[KW]           Alpha_Beta

```



```
SEQ  MAACRALKAVLVDSLGLHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERL
PRD  cccccccccccccccccccccchhhhhhhhhccccccccccccchhhhhhhh

SEQ  RKLEFDISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTSDPNAVVMGLAPE
PRD  hhhccccccccceehhhhhhhhhccccccccchhhhhcccccccccccccccccccc

SEQ  HFHYQILNQAFRLLLDGAPLIAIHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKP
PRD  chhhhhhhhhhhccccccccccccccccccccccccchhhhhhhcccccccccccc

SEQ  EKTFEALRGTCPEPEEAVMIGDDCRDDVGGAQDVGMLGILVKTGKYRASDEEKINPPP
PRD  cchhhhhhhhhccccccccccccchhhhhhhhhcccccccccccccccccccccccc

SEQ  YLTCEFPFAVDHILQHL
PRD  cccccchhhhhhhhhccc
```

(No Prosite data available for DKFZphfbr2_78d13.2)

(No Pfam data available for DKFZphfbr2_78d13.2)

DKFZphfbr2_78k24

group: metabolism

DKFZphfbr2_78k24 encodes a novel 372 amino acid protein with similarity to *Mus musculus* ubiquitin specific protease UBP43.

The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The new protein can find application in modulation of protein stability/degradation in cells.

Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.

strong similarity to mouse ubiquitin specific protease UBP43

Sequenced by MediGenomix

Locus: unknown

Insert length: 1874 bp

Poly A stretch at pos. 1852, polyadenylation signal at pos. 1836

```
1 AGTCCCGACG TGGAACTCAG CAGCGGAGGC TGGACGCTTG CATGGCGCTT
51 GAGAGATTCC ATCGTGCCTG GCTCACATAA GCGCTTCCTG GAAAGTGAAGT
101 CGTGCTGTCC TGAACGCGGG CCAGGCAGCT GCGGCCTGGG GGTTTTGGAG
151 TGATCACGAA TGAGCAAGGC GTTTGGGCTC CTGAGGCAAA TCTGTCAGTC
201 CATCCTGGCT GAGTCCTCGC AGTCCCGGGC AGATCTTGAA GAAAAGAAAGG
251 AAGAAGACAG CAACATGAAG AGAGAGCAGC CCAGAGAGCC TCCCAGGGCC
301 TGGGACTACC CTCATGGCCT GGTGGTTTA CACAACATTG GACAGACCTG
351 CTGCCCTTAAC TCCTTGATTG AGGTGTTCGT AATGAATGTG GACTTCACCA
401 GGATATTGAA GAGGATCAGC GTGCCAGGG GAGCTGACGA GCAGAGGAGA
451 AGCGTCCCTT TCCAGATGCT TCTGCTGCTG GAGAAGATGC AGGACAGCCG
501 GCAGAAAGCA GTGCGGCCCC TGGAGCTGGC CTACTGCCTG CAGAAGTGCA
551 ACGTGCCCTT GTTTGTCCAA CATGATGCTG CCCAACTGTA CCTCAAACCTC
601 TGGAACTTGA TTAAGGACCA GATCACTGAT GTGCACTTGG TGGAGAGACT
651 GCAGGCCCTG TATACGATCC GGTGAAGGA CTCCTTGATT TGCCTTGACT
701 GTGCCATGGA GAGTAGCAGA AACAGCAGCA TGCTCACCTT CCCACTTTCT
751 CTTTTTGATG TGGACTCAAA GCCCTGAAG ACCTGGAGG ACGCCCTGCA
801 CTGCTTCTTC CAGCCAGGG AGTTATCAAG CAAAAGCAAG TGCTTCTGTG
851 AGAACTGTGG GAAGAAGACC CGTGGGAAAC AGGTCTTGAA GCTGACCCAT
901 TTGCCCCAGA CCCTGACAAT CCACCTCATG CGATTCTCCA TCAGGAATTC
951 ACAGACGAGA AAGATCTGCC ACTCCCTGTA CTTCCCCAG AGCTTGGATT
1001 TCAGCCAGAT CCTTCCAATG AAGCGAGAGT CTTGTGATGC TGAGGAGCAG
1051 TCTGGAGGGC AGTATGAGCT TTTTGTGTG ATTGCGCAGC TGGGAATGGC
1101 AGACTCCGGT CATTACTGTG TCTACATCCG GAATGCTGTG GATGGAAAAT
1151 GGTTCGTGCT CAATGACTCC AATATTGCTT TGGTGTCTG GGAAGACATC
1201 CAGTGTACCT ACGGAAATCC TAACTACCAC TGGCAGGAAA CTGCATATCT
1251 TCTGTTTTAC ATGAAGATGG AGTGCTAATG GAAATGCCCA AAACCTTCAG
1301 AGATTGACAC GCTGTCAATT TCCATTCCG TTCCTGGATC TACGGAGTCT
1351 TCTAAGAGAT TTTGCAATGA GGAGAAGCAT TGTTTTCAAA CTATATAACT
1401 GAGCCTTATT TATAATTAGG GATATTATCA AAATATGTAA CCATGAGGCC
1451 CCTCAGGTCC TGATCAGTCA GAATGGATGC TTTCAACAGC AGACCCGGCC
1501 ATGTGGCTGC TCGGTCCTGG GTGCTCGCTG CTGTGCAAGA CATTAGCCCT
1551 TTAGTTATGA GCCTGTGGGA ACTTCAGGGG TTCCCAGTGG GGAGAGCAGT
1601 GGCAGTGGGA GGCATCTGGG GGCCAAAGGT CAGTGGCAGG GGGTATTTC
1651 GTATTATACA ACTGCTGTGA CCAGACTTGT ATACTGGCTG AATATCAGTG
1701 CTGTTTGTA TTTTCACTT TGAGAACCAA CATTAATCC ATATGAATCA
1751 AGTGTTTTGT AACTGCTATT CATTTATTCA GCAAATATTT ATTGATCATC
1801 TCTTCTCCAT AAGATAGTGT GATAAACACA GTCATGAATA AAGTTATTTT
1851 CCACAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry AC005500 from database EMBL:

, complete sequence.

Score = 859, P = 5.7e-143, identities = 175/179

8 exons matching Bp 317-1230

Medline entries

99182491:

A novel ubiquitin-specific protease, UBP43, cloned from leukemia fusion protein AML1-ETO-expressing mice, functions in hematopoietic cell differentiation.

Peptide information for frame 1

ORF from 160 bp to 1275 bp; peptide length: 372
 Category: strong similarity to known protein
 Classification: Protein management
 Prosite motifs: UCH_2_2 (302-320)

```

1 MSKAFGLLRQ ICQSILAESS QSPADLEEK EEDSNMKREQ PRERPRAWDY
51 PHGLVGLHNI GQTCCNLNLI QVFVMNVDFTRILKRITVPR GADEQRRSVP
101 FQMLLLEKMQDSRQKAVRP LELAYCLQKC NVPLFVQHDA AQLYLKLWNL
151 IKDQITDVHL VERLQALYTI RVKDSLICVD CAMESSRNSS MTLPLSLFD
201 VDSKPLKTLE DALHCFQFPR ELSSSKSCFC ENCGKKTRGK QVLKLTHLPQ
251 TLTIHLMRFS IRNSQTRKIC HSLYFPQSLD FSQILPMKRE SDAEEQSGG
301 QYELFAVIAH VGMADSGHYC VYIRNAVDGK WFCFNDNIC LVSWEDIQCT
351 YGNPNYHWQE TAYLLVYMK EC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78k24, frame 1

TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds., N = 1, Score = 1367, P = 1e-139

SWISSPROT:UBPE_DROME UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)., N = 2, Score = 248, P = 5.3e-33

>TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds.
 Length = 368

HSPs:

Score = 1367 (205.1 bits), Expect = 1.0e-139, P = 1.0e-139
 Identities = 262/369 (71%), Positives = 295/369 (79%)

```

Query:      1 MSKAFGLLRQICQSILAESSQSPADLEEK EEDSNMKREQPRERPRAWDYPHGLVGLHNI 60
            M K FGLLR+ CQS++AE Q A LEE E KR R+ AWD PHGLVGLHNI
Sbjct:      1 MGKGFGLLRKPCQSVVAEPQQYSA-LEE--ERTMKRKRVLSDLCSAWDSPHGLVGLHNI 57

Query:      61 GQTCCNLNLIQVFVMNVDFTRILKRITVPRGADEQRRSVPFQMLLLEKMQDSRQKAVRP 120
            GQTCCNLNLIQVF+MN+DF ILKRITVPR A+E++RSVPFQ+LLLLEKMQDSRQKA+ P
Sbjct:      58 GQTCCNLNLIQVFMMNMDFRMILKRITVPRSAEERKRSVPFQLLLLEKMQDSRQKALLP 117

Query:      121 LELAYCLQKCNVPLFVQHDAQAQLYLKLWNLIKDQITDVHLVERLQALYTIIRVKDSLICVD 180
            EL CLQK NVPLFVQHDAQAQLYL +WNL KDQITD L ERLQ L+TI ++SLICV
Sbjct:      118 TELVQCLOKYNVPLFVQHDAQAQLYLTIWNLTKDQITDTERLQGLFTIWTQESLICVG 177

Query:      181 CAMESSRNSSMLTLPPLSLFDVDSKPLKTLEDALHCFQFRELSSSKSCFCENCGKKTRGK 240
            C ESSR S +LTL L LFD D+KPLKTLEDAL CF QP+EL+S C CE CG+KT K
Sbjct:      178 CTAESSRRSKLLTSLPLFDKDAKPLKTLEDALRCFVQPKELASSDMC-CETCGEKTPTWK 236

Query:      241 QVLKLTHLPQTLTIHLMRFSIRNSQTRKICHSLYFPQSLDQSLPMKRESCDAEEQSGG 300
            QVLKLTHLPQTLTIHLMRFS RNS+T KICHS+ FPQSLDQSLQ+LP + + D +EQS
Sbjct:      237 QVLKLTHLPQTLTIHLMRFSARNSRTEKICHSVNFPQSLDQSLQVLPTEEDLGDTKEQSEI 296

Query:      301 QYELFAVIAHVGMADSGHYCVYIRNAVDGKWFCFNDNICLVSWEDIQCTYGNPNYHWQE 360
            YELFAVIAHVGMAD GHYC YIRN VDGKWFCFNDN++C V+W+D+QCTYGN Y W+E
Sbjct:      297 HYELFAVIAHVGMADFGHYCAYIRNPVDGKWFCFNDN+HVCWVTWKDVQCTYGNHRYRWRE 356

Query:      361 TAYLLVYMK 369

```



```
HMM_NAME      Ubiquitin carboxyl-terminal hydrolases family 2
HMM            *YdLYgVICHYGntldyGHYWaYVKNenhHRWkWYYFDEtV*
               Y+L++VI H G   D+GHY +Y++N   ++KW++F+D+++
Query          302 YELFAVIAHVG-MADSGHYCVYIRNAV--DGKWFCFNDNSI   339
```


DKFZphfbr2_78n23

group: brain derived

DKFZphfbr2_78n23 encodes a novel 329 amino acid protein with similarity to A.thaliana F26P21.80 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F26P21.80

Sequenced by MediGenomix

Locus: /map="89.1 cR from top of Chr19 linkage group"

Insert length: 1447 bp

Poly A stretch at pos. 1374, polyadenylation signal at pos. 1353

```

1  TACAACTTCC GGCTGTAAAG ATGGCGGCTT CCTAGTGAGT CGGCGGCTGA
51  CTTAGAAGGA GGTTCAGGCT ACGGTGAGCC GAAGCCACAC AGGAGCCATG
101 GAAGTGGCAG AGCCCAGCAG CCCCACTGAA GAGGAGGAGG AGGAAGAGGA
151 GCACTCGGCA GAGCCTCGGC CCCGCACTCG CTCCAATCCT GAAGGGGCTG
201 AGGACCGGGC AGTAGGGGCA CAGGCCAGCG TGGGCAGCCG CAGCGAGGGT
251 GAGGGTGAGG CCGCCAGTGC TGATGATGGG AGCCTCAACA CTTCAGGAGC
301 CGGCCCTAAG TCCTGGCAGG TGCCCCCGCC AGCCCCTGAG GTCCAAATTC
351 GGACACCAAG GGTCAACTGT CCAGAGAAAG TGATTATCTG CCTGGACCTG
401 TCAGAGGAAA TGTCAC TGCC AAAGCTGGAG TCGTTCAACG GCTCCAAAAC
451 CAACGCCCTC AATGTCTCTC AGAAGATGAT TGAGATGTTC GTGCGGACAA
501 AACACAAGAT CGACAAAAGC CACGAGTTTG CACTGGTGGT GGTGAACGAT
551 GACACGGCCT GGCTGTCTGG CCTGACCTCC GACCCCGCGG AGCTCTGTAG
601 CTGCCTCTAT GATCTGGAGA CGGCCTCCTG TTCCACCTTC AATCTGGAAG
651 GACTTTTCAG CCTCATCCAG CAGAAAAC TGAGTTCCGGT CACAGAGAAC
701 GTGCAGACGA TTCCCCCGCC ATATGTGGTC CGCACCATCC TTGTCTACAG
751 CGGTCCACCT TGCCAGCCCC AGTTCTCCTT GACGGAGCCC ATGAAGAAAA
801 TGTTCCAGTG CCCATATTTT TTCTTTGACG TTGTTTACAT CCACAATGGC
851 ACTGAGGAGA AGGAGGAGGA GATGAGTTGG AAGGATATGT TTGCCTTCAT
901 GGGCAGCCTG GATACCAAGG GTACCACTA CAAGTATGAG GTGGCACTGG
951 CTGGGCCAGC CCTGGAGTTG CACAAC TGCGGAAACT GTTGGCCAC
1001 CCCCTGCAGC GGCCTTGCCA GAGCCATGCT TCCTACAGCC TGCTGGAGGA
1051 GGAGGATGAA GCCATTGAGG TTGAGGCCAC TGCTGAACC ATCCCTGTAC
1101 ATCTGCACCT TCTTGTCGAA GGAAGTCCTT GGCTAAAGC CTTGGTTCTC
1151 AAAGTGGGTT CCTTGGGACC TCCGGGGTGG GGGGGTTCCA GGAGGCACGT
1201 AGGGTACCTT GCAGGGTCCT AGGAGGGAAA CCCAGGATTC CAGGAGGGAT
1251 CCCAGGAAC GTGGGCACCC ATTTTCTGTG TCTCCAGCC CATTTCCACT
1301 CCTAGTTTGT CATGGATAAT TTTTGTCTT CCCTGTGTGA TTTTGGCCAT
1351 CAAAATAAAA ATTTGAGACT CGTTAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAAGAAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HS806352 from database EMBL:
human STS EST192543.
Score = 1285, P = 2.5e-51, identities = 263/266

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 98 bp to 1084 bp; peptide length: 329
Category: similarity to unknown protein
Classification: no clue

1 MEVAEPSSPT EEEEEEEHS AEPRPRTRSN PEGAEDRAVG AQASVGSRSSE


```

51 GEGEASADD GSLNTSGAGP KSWQVPPAP EVQIRTPRVN CPEKVIICLD
101 LSEEMSLPKL ESFNGSKTNA LNVSQKMIEM FVRTKHKIDK SHEFALVVVN
151 DDTAWLSGLT SDPRELCSCL YDLETASCST FNLEGLFSLI QQKTELPVTE
201 NVQTIPPPYV VRTILVYSRP PCQPQFSLTE PMKKMFQCPY FFFDVVYIHN
251 GTEEEKEEMS WKDMFAFMGS LDTKGTSYKY EVALAGPALE LHNCMAKLLA
301 HPLQRPCQSH ASYSLLEED EAIEVEATV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78n23, frame 2

PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana, N = 1, Score = 142, P = 1.5e-07

>PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana
Length = 264

HSPs:

Score = 142 (21.3 bits), Expect = 1.5e-07, P = 1.5e-07
Identities = 56/216 (25%), Positives = 97/216 (44%)

```

Query:  93 EKVIICLDL-SEEMSLPKLESFNGSKTNALNVSQKMIEMFVRTKHKIDKSHEFALVVVND 151
      E ++IC+D+ +E M K NG + ++ I +F+ K I+ H FA +
Sbjct:  26 EDILICIDVDAESMVEMKTTGTNGRPLIRMECVKQAILFIHNKLSINPDHRAFATLAK 85

Query:  152 DTAWLSG-LTSDPRELCSCLYDLE-TASCSTFNLEGLFSLIQQKTELPVTENVQTIPPPY 209
      AWL TSD + L L S S +L LF Q+ ++ +N
Sbjct:  86 SAAWLKKEFTSDAESAVASLRGLSGNKSSSRADLTLLFRAAAQEAQVSRQN-----R 138

Query:  210 VVRTILVYSRPPCQPQFSLTEPMKKMFQCPYFFFDVVYIHNGTEEEKEEMSWKDMF-AFM 268
      + R IL+Y R +P P+ + F DV+Y+H ++ + +D++ + +
Sbjct:  139 IFRVILIYCRSSMRPTEHW--PLNQKL----FTLDVMYLH---DKPSPDNCPCQDVYDSL 189

Query:  269 GSLD--TKGTSYKYEVALAGPALELHNCMAKLLAHPLQRPCQ 308
      +++ ++ Y +E G A + M+ LL HP QR Q
Sbjct:  190 DAVEHVSEYEGYIFESG-QGLARSVFKPMSMLLTHPQORCAQ 230

```

Pedant information for DKFZphfbr2_78n23, frame 2

Report for DKFZphfbr2_78n23.2

```

[LENGTH] 329
[MW] 36560.10
[pI] 4.60
[HOMOL] PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana 7e-07
[KW] Alpha Beta
[KW] LOW_COMPLEXITY 9.73 %

```

```

SEQ MEVAEPSSPTEEEEEEHSAEPRPRTSRNPEGAEDRAVGAQASVGSRSEGEASADD
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhcccccccccccccc

```

```

SEQ GSLNTSGAGPKSWQVPPAPAEVQIRTPRVNCPEKVIICLDLSEEMSLPKLESFNGSKTNA
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ LNVSQKMIEMFVRTKHKIDKSHEFALVVVNDDTAWLSGLTSDPRELCSCLYDLETASCST
SEG .....
PRD ehhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhcccccccccc

```

```

SEQ FNLEGLFSLIQQKTELPVTENVQTIPPPYVVRTILVYSRPPCQPQFSLTEPMKKMFQCPY
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhheeee

```

```

SEQ FFFDVVYIHNGTEEEKEEMSWKDMFAFMGSLDTKGTSYKYEVALAGPALELHNCMAKLLA
SEG .....
PRD eeeeeeeccccchhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhh

```

```

SEQ HPLQRPCQSHASYSLLEEDAEIEVEATV
SEG .....xxxxxxxxxxxxx...
PRD hccccccccchhhhhhhhhhhhhhhcccc

```


(No Prosite data available for DKFZphfbr2_78n23.2)

(No Pfam data available for DKFZphfbr2_78n23.2)

DKF2phfbr2_7a24

group: brain derived

DKF2phfbr2_7a24 encodes a novel 142 amino acid protein with similarity to the C-terminal part of transforming growth factor-beta activated kinases.

The novel protein shows only similarity to the C-terminus of such kinases; no kinase domain is present.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C-terminus of TGF-beta-activated kinase

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1697 bp

No poly A stretch found, no polyadenylation signal found

```
1 GGGGAGAGAG GGGTTGTGAA GGGGAAGCGGA AGGGAAGGGA AGGGAGGTCC
51 CGTGGGACGC TGGGGTCTGG GGTAGAGCAG GTAGCAGCGT GCTGCCCTGA
101 CAGCTGTCTC CGCTCCTCAG ATTGTCACTG GCTGCTATGC AGCAGGTGCA
151 GCCTGGTCTC TCACTGAGTC TCTACTCCAC AAAGGCAACG ACTGGCCAAG
201 GCAGTGGCTG GCTCTGGGTT ACACAAGTGC AGACACTCAA CTAAGTGAGC
251 TGGGAAGACCC AGGAGAAGGC GGAGGCTCAG GTGCCACAT GATCAGCACA
301 GCCAGGGTAC CTGCTGACAA GCCTGTACGC ATCGCCTTTA GCCTCAATGA
351 CGCCTCAGAT GATACACCCC CTGAAGACTC CATTCCCTTG GTCTTCCAG
401 AATTAGACCA GCAGCTACAG CCCCTGCCGC CTTGTCATGA CTCCGAGGAA
451 TCCATGGAGG TGTTTCAGACA GCACTGCCAA ATAGCAGAAG AATACCTTGA
501 GGTCAAAAAG GAAATCACCC TGCTTGAGCA AAGGAAGAAG GAGCTCATTG
551 CCAAGTTAGA TCAGGCAGAA GAGGAGAAGG TGGATGCTGC TGAGCTGGTT
601 CGGGAATTCT AGGCTCTGAC GGAGGAGAA CCGACGTTGA GGTGGCCCA
651 GTCTCAATGT GTGGAACAAC TGGAGAACT TCGAATACAG TATCAGAAGA
701 GGCAGGGCTC GTCCTAACTT TAAATTTTTC AGTGTGAGCA TACGAGGCTG
751 ATGACTGCCC TGTGCTGGCC AAAAGATTTT TATTTTAAAT GAATAGTGAG
801 TCAGATCTAT TGCTTCTCTG TATTACCCAC ATGACAACCTG TCTATAATGA
851 GTTTACTGCT TGCCAGCTTC TAGCTTGAGA GAAGGGATAT TTTAAATGAG
901 ATCATTAAAC TGAAACTATT ACTAGTATAT GTTTTGGAG ATCAGAATTC
951 TTTTCCAAAG ATATATGTTT TTTTCTTTT TAGGAAGATA TGATCATGCT
1001 GTACAACAGG GTAGAAAATG GTAAAAATAG ACTATTGACT GACCCAGCTA
1051 AGAATCGCGG GCTGAGCAGA GTTAAACCAT GGGACAAACC CATAACATGT
1101 TCACCATAGT TTCACGTATG TGTATTTTAA AATTTCATGC CTTTAAATAT
1151 TCAAATATGC TCAAATTTAA ACTGTCAGAA ACTTCTCTGC ATGTATTTAT
1201 ATTTGCCAGA GTATAAACTT TTATACTCTG ATTTTATCC TTCAATGATT
1251 GATTATACTA AGAATAAATG GTCACATATC CTAAAAGCTT CTTTCATGAA
1301 TTATTAGCAG AAACCATGTT TGAACCAAA GCACATTGTC CAATGCTAAC
1351 TGGCTGTTGT AATAATAAAC AGATAAGGCT GCATTTGCTT CATGCCATGT
1401 GACCTCACAG TAAACATCTC TGCCTTTGCC TGTGTGTGTT CTGGGGGAGG
1451 GGGGACATGG AAAAATATTG TTTGGACATT ACTTGGGTGA GTGCCCATGA
1501 AGACATCAGT GAACTTGTA CTATTGTTT GTTTTGGATT TAAGGAGATG
1551 TTTTAGATCA GTAACAGCTA ATAGGAATAT GCGAGTAAAT TCAGAATTGA
1601 AACAAATTTCT CTTGTTCTA CCTATCACCA CATTCTCTCA AATTGAATCT
1651 TTTGTTATAT GTCCATTCTT ATTCATGTAA CTTCTTTTTC ATTAAC
```

BLAST Results

No BLAST result

Medline entries

98130593:
Role of TAK1 and TAB1 in BMP signaling in early Xenopus development.

Peptide information for frame 1

ORF from 289 bp to 714 bp; peptide length: 142
Category: similarity to known protein

1 MISTARVPAD KPVRIAFSLN DASDDTPPED SIPLVFPELD QQLQPLPPCH
51 DSEESMEVFR QHCQIAEEYL EVKKEITLLE QKKELIAKL DQAEEEKVDA
101 AELVREFEAL TEENRTLRLA QSQCVEQLEK LRIQYQKRQG SS

BLASTP hits

Entry U92030_1 from database TREMBL:
product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1 mRNA,
complete cds.
Score = 343, P = 1.3e-30, identities = 69/143, positives = 104/143

Entry AB009356_1 from database TREMBL:
product: "TGF-beta activated kinase 1a"; Homo sapiens mRNA for
TGF-beta activated kinase 1a, complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry MMPK_1 from database TREMBL:
product: "TAK1 (TGF-beta-activated kinase)"; Mouse mRNA for TAK1
(TGF-beta-activated kinase), complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry AB009357_1 from database TREMBL:
product: "TGF-beta activated kinase 1b"; Homo sapiens mRNA for
TGF-beta activated kinase 1b, complete cds.
Score = 339, P = 3.2e-30, identities = 67/143, positives = 104/143

Entry AB009358_1 from database TREMBL:
product: "TGF-beta activated kinase 1c"; Homo sapiens mRNA for
TGF-beta activated kinase 1c, complete cds.
Score = 144, P = 3.8e-09, identities = 30/67, positives = 47/67

Alert BLASTP hits for DKFZphfbr2_7a24, frame 1

PIR:JC5955 transforming growth factor-beta activated kinase (EC
-.-.-) 1a - Human, N = 1, Score = 339, P = 3e-30

>PIR:JC5955 transforming growth factor-beta activated kinase (EC -.-.-) 1a
- Human

Length = 579

HSPs:

Score = 339 (50.9 bits), Expect = 3.0e-30, P = 3.0e-30
Identities = 67/143 (46%), Positives = 104/143 (72%)

Query: 1 MISTARVPADKPVRI-AFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVF 59
MI+T+ ++KP R ++ +D+D ++SIP+ + LD QLQPL PC +S+ESM VF
Sbjct: 437 MITTSGPTSEKPTRSHPWTPDDSTDTNGSDNSIPMAYLTLDHQLQPLAPCPNSKESMAVF 496
Query: 60 RQHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRL 119
QHC++A+EY++V+ EI LL QRK+EL+A+LDQ E+++ + + LV+E + L +EN++L
Sbjct: 497 EQHCKMAQEYMKVQTEIALLLQRKQELVAELDQDEKQNTSRLVQEHKKLLDENKSLST 556
Query: 120 AQSQCVEQLEKLRIQYQKRQGSS 142
QC +QLE +R Q QKRQG+S
Sbjct: 557 YYQQCKKQLEVIRSQQQKRQGTS 579

Pedant information for DKFZphfbr2_7a24, frame 1

Report for DKFZphfbr2_7a24.1

[LENGTH] 142
[MW] 16377.53
[pI] 4.64
[HOMOL] TREMBL:U92030_1 product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1
mRNA, complete cds. 6e-26
[PROSITE] CK2_PHOSPHO_SITE 3


```

[PROSITE]      PKC_PHOSPHO_SITE      2
[PROSITE]      ASN_GLYCOSYLATION     1
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY         7.04 %
[KW]           COILED_COIL           33.10 %

```

```

SEQ      MISTARVPADKPVRIAFSLNDASDDTTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVFR
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS     .....

```

```

SEQ      QHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRLA
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS     ...CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ      QSQCVEQLEKLRIQYQKRQGS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS     .....

```

Prosite for DKFZphfbr2_7a24.1

PS00001	114->118	ASN_GLYCOSYLATION	PDOC00001
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00006	18->22	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006

Pfam for DKFZphfbr2_7a24.1

```

HMM_NAME      TNFR/NGFR cysteine-rich region
HMM            *CpeGtYtDWNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*
               C++ + + + +Q C++ E+ ++++++ T + ++
Query         49  CHDSEESMEVF-RQH--CQIAEE--YLEVKKEITLLEQRKK      84

```


DKFZphfbr2_7e22

group: brain derived

DKFZphfbr2_7e22.2 encodes a novel 286 amino acid protein similar to b561 cytochromes

The new protein shows strong similarity to B561 cytochromes, but contains no heme binding site. In addition, a myc-type, helix-loop-helix dimerization domain is present. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to cytochrome b561

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 4254 bp

Poly A stretch at pos. 4234, polyadenylation signal at pos. 4217

```

1 GGGGACTACC CAGAGGGCTG CCGCCGCCCTC TCCAAGTTCT TGTGGCCCCC
51 GCGGTGCGGA GTATGGGCGG CTGATGGCCA TGGAGGGCTA CCGGCGCTTC
101 CTGGCGCTGC TGGGGTCGGC ACTGCTCGTC GGCTTCCTGT CGGTGATCTT
151 CGCCCTCGTC TGGGTCTCC ACTACCGAGA GGGGCTTGGC TGGGATGGGA
201 GCGCACTAGA GTTAACTGG CACCCAGTGC TCATGGTCAC CGGCTTCGTC
251 TTCATCCAGG GCATCGCCAT CATCGTCTAC AGACTGCCGT GGACCTGGAA
301 ATGCAGCAAG CTCCTGATGA AATCCATCCA TGCAGGGTTA AATGCAGTTG
351 CTGCCATTCT TGCAATTATC TCTGTGGTGG CCGTGTTTGA GAACCACAAT
401 GTTAAACAATA TAGCCAATAT GTACAGTCTG CACAGCTGGG TTGGACTGAT
451 AGCTGTGATA TGCTATTTGT TACAGCTTCT TTCAGGTTTT TCAGTCTTTC
501 TGCTTCCATG GGCTCCGCTT TCTCTCCGAG CATTTCTCAT GCCCATACAT
551 GTTTATTCTG GAATTGTGAT CTTTGGAACA GTGATTGCAA CAGCACTTAT
601 GGGATTGACA GAGAAACTGA TTTTTCCTT GAGAGATCCT GCATACAGTA
651 CATTCCCGCC AGAAGGTGTT TTCGTAAATA CGCTTGGCCT TCTGATCCTG
701 GTGTTCGGGG CCCTCATTTT TTGGATAGTC ACCAGACCGC AATGGAAACG
751 TCCTAAGGAG CCAAATTCTA CCATTCTTCA TCCAATGGA GGCACCTGAAC
801 AGGGAGCAAG AGGTTCCATG CCAGCCTACT CTGGCAACAA CATGGACAAA
851 TCAGATTCAG AGTTAAACAA TGAAGTAGCA GCAAGGAAAA GAAACTTAGC
901 TCTGGATGAG GCTGGGCAGA GATCTACCAT GTAAATGTT CTAGAGATAG
951 AGCCATATAA CGTCACGTTT CAAACTAGC TCTACAGTTT TGCTTCTCCT
1001 ATTAGCCATA TGATAATTGG GCTATGTAGT ATCAATATTT ACTTTAATCA
1051 CAAAGGATGG TTTCTTGAAA TAATTGTAT TGATTGAGGC CTATGAACCTG
1101 ACCTGAATTG GAAAGGATGT GATTAATATA AATAATAGCA GATATAAATT
1151 GTGGTTATGT TACCTTTATC TTGTTGAGGA CCACAACATT AGCACGGTGC
1201 CTTTGTGCAGA ATAGATACTC AATATGTGAA TATGTGCTA CTAGTAGTTA
1251 ATTGATATAA CTGGCAGCAT CCCTGGCCTG TTGTCATGCA GTCATTTCCT
1301 GTTAATTCTG GGAGACAATG ATTTCAACA TAGAGGGAAG CAGTCCTAAA
1351 AGTTTAAAT CCGATAAGGA ATATCTGGGA CAGGGTTTAG ATCATGACTC
1401 TACACAGATA CCATGATGAG AGTATATTAA AGAAATTTAG GAAAGCACCT
1451 GGTTCCTTTC TCCCATGCC TGCCTTCTGC TCCCTCCCA GCTGGTTTGG
1501 GCTCAAATTG TCCCTGGAGA CTAGGGTTTA TGTAGGGTA TTGATAGATT
1551 AGAGCAGGTG GTTGAAGAGA TCTTCTCTGG TCAGACTTGG AAGAATTTCC
1601 AAAAGTGAAG TTAGCCCCAA GACTTCCCTA GGGTTGATGT ACTTTATGAT
1651 CCAGATGCTA AACTTCTTAG AATGAAAATA TGCTTCAACA CTTAAGTAGC
1701 ATACACTGCC CTACAAACCT CAGAGAGCAC TTTTCCCCAA GTTCTTGTTT
1751 TTATTTTGA AAGTACTCAC ACAGCACTTA CTATGCTCCA AACACTCCTC
1801 TAAGCACTTT ACACATATTA GCTCATTGAG TCCCCAGACA GACGGGATGA
1851 AGTAGGTATT GTTACTGTTT CCATTTTACA GGTGAGAGAT TTGAAGCCTG
1901 GGGAGGCTAG TAACTCACCC CAAGGTCACA CGGCTCATA ATGGTGGGAC
1951 TGAGACTCAG ATGCAGGCAG TCTGGCACCT CAGTCTGGAT TCTAACCAAT
2001 TCACTAAGCT ATTTTGTGCT TGTACTACTT TGACCCACCC CTGAATAAAC
2051 CTCATTTGCT GGAGTGGGGT GTAGTTATTA AAGGGATGCT TTTTACCTTT
2101 TGCTGCTGTC TGTGGCAGAT TCCCCAGATA ACCAAGGAAA AGGGGCCACC
2151 CATACCTGGA AATAGGCCAT AGGGCCCCTA CTACTGCCAA CAAGCCATGG
2201 CCTACCTTGA CACTTGTGTT ATCTTAAAT TGTTCTTGG TAACAAAAGA
2251 TTTGGACAGG CATATCTGTA GCTTTCAGT TAATTAATTG CAATATTTT
2301 TTCTTCAGGA TTTTAGCTGC TGAACAACCT TCAGTTTGGG GCTAAAAGAG
2351 ACCGTGCTCA TGGTCTGCCC TCCCCTGGGG CAATAGCTAG GGTCTTTCCT
2401 GATTTTATG GAATTTTAGG GGATATTTTG AGCTTTGGGT TCTCAGTAGT

```



```

2451 GAATTGAGAC TTGGAGGTGA CTTTTCATGT TTGGAGTATC ATCTCTGTCT
2501 GGGCTCTGGG CTGACAAATT AAAACCTAGA GTAGTGCTTA TGCTGAAATG
2551 ATACTTTTCA TTTTGTGGTT GATTTTITG CCTTCCCTTC AATTTTAAAC
2601 TGAAGCATT TAATGTGGGT AGAACTCTA CACCAATAC ACTAAACATT
2651 TTGGTGCTTA GTGGATTTCT TTTTAGGTAA CTGGTACTTA CTTCACAAAG
2701 CTGAATACAA GCCCACTCC ATCATATCCC TTAAACTTCA TGAAAAACCA
2751 TTCAAGATCC CCTTGCTGCA ACACGTCTCT CTCTCTCTCT ACTAAATTCT
2801 ATTTCCAAAA TTGGTAATAG AGCCAGAAGG ATCCCCAGTA CCCAGCCCTC
2851 TGCCTGGCAC AAAGTGGTAG CACAATTAAA TTCAGTATGG GTGGAGCATG
2901 GTACAGTCTT GGTGCCATAG AAGGAGTAGT TGCATAGTCA CACATCATTT
2951 GATAAGTTGG ATGTTCCATT ACATAGAGGA ACACAAATTT CCAGGGTTTT
3001 TGGAGGAAGG GATTAGATAG CGACTAAGCC GCCAGAATTG AGGTGGCCAT
3051 TCCTTTTGT ATAGGCTAAG AAACAGGTTA TCAGTGAAAA GTTAATTATG
3101 GCTTTGGCAC TAGAATAGCA CTGTTGCAAA GTATTTAAGC ACCCCCCATC
3151 TCAGCCCTTT ATTTTATCTT TCATGTGGGC TAATGTGAGG ATAATCTTAC
3201 AGATATTATA GGAATTTCTT TTCTATCTTT ATGAAAACAA CGTATATAAA
3251 ATATATCTAG AAAACCTTTG TTTGAGACTC TTATTTAATG GGCTTTTGAT
3301 TCTAATGATA ATTGTACCTT TATCTTTCAA AAGCTGATAT TTCCTACCTA
3351 AGCATCTCCC GAGAAAAATA TCTCATTAAA AAGCCCATAA ATAATAGGGG
3401 AGAAGAAAGC CTTAGGTATC AATTCCAAAA CAGTGATTGA AATTTCCCAA
3451 AATAATTATG GCTTCTGTCA TCTCCAGAGA TAATCTGGCT TGGTTTACCC
3501 CATAATCTAA TTTGAGAAAA GAAAGCTTTA TTTTAACACT CATCTGAATC
3551 AACATTAAAG CCTTTTCTCT CAAAGCGTTT ATTGAGAAAC TCAAATGAAT
3601 ATACTTTTTG AATTACTGTC ATCAAAAGTG TACGGCTTCC TGTGCTGCTT
3651 GTGTCAAATG GAACCTGCCC TCTAAAGCAC TTTCTTTCTT TTACTTGCGT
3701 GGTTCATGT AAGCTGTGCT GTTTAGAAAC AACATCTCAG ACTTTACAAA
3751 GAAATGACAA AGAAGGCAAT TGCACTTTTT AAGGGATATC GACAAGCAGT
3801 TTCTGTTTTT TAAAGGACAA AATACAGAGT GTGTGTCATT TTTAATTAGA
3851 TTCTTTCCCC TGCTGAGTTG GAAATCCAG TGCAGCACTG ATTGACCACA
3901 GTTGCCAATC TAAAGCACA AAGACAGAAG TAAAGCTTTA TGCTAATTTT
3951 ATTTCAATAT GATAGAAAAT TTATCTTGGT ATGTCCTTTT TTAGATAACT
4001 CCAGCAGGAA ACTGTAAGTG CTATGTCTTT AGGAAAACGT AGAAGAAAGA
4051 ACATTATTAT TCTTTAATTC CTACAAGGTA CTTGAAAACC TTAAGTGAAA
4101 AAGATTCTTA TCTTTTATC TTGGCGCATT TATGGAAAAA ATATTAACGT
4151 TCCTGAATAT TTTATAATTT TGTAGGAAAA ATATGCATCT ATTTTCTCTT
4201 GACTTCTTTT ATATAGTAAT AAAAGTTATT TTGGAAAAAA AAAAAAAAAA
4251 AAAA

```

BLAST Results

Entry HSG20626 from database EMBL:
human STS A005227.
Score = 860, P = 3.0e-32, identities = 176/181

Medline entries

89030633:
The structure of cytochrome b561, a secretory vesicle-specific electron transport protein.

Peptide information for frame 2

ORF from 74 bp to 931 bp; peptide length: 286
Category: strong similarity to known protein
Classification: unset

```

1 MAMEGYRRFL ALLGSALLVG FLSVIFALVW VLHYREGLGW DGSALFNWH
51 PVLMTVGTFV IQGIAIIVYR LPWTWKCSKL LMKSIHAGLN AVAAILAIIS
101 VVAVFENHNH NNIANMYSLH SWVGLIAVIC YLLQLLSGFS VFLLPWAPLS
151 LRAFLMPIHV YSGIVIFGTV IATALMGLTE KLIFSLRDPY YSTFPPEGVF
201 VNTLGLLILV FGALIFIVT RPQWKRPEP NSTILHPNGG TEQGARGSMF
251 AYSGNMMDKS DSELNNEVAA RKRNLALDEA GQRSTM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_7e22, frame 2

SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score

PIR:S01167 cytochrome b561 - bovine, N = 1, Score = 457, P = 2.7e-43
 SWISSPROT:C561_PIG CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score = 452, P = 9.1e-43
 PIR:S53321 cytochrome B561 - human, N = 1, Score = 451, P = 1.2e-42

HSPs :

```

Query:      18 LVGFLSVIFALVWVLHYREGLGWDGSALEFNWHPVLMVTGTFVFIQIGIAIIVYRPLWTKWC 77
              L+G  V    W+  YR G+ W+ SAL+FN HP+ MV G VF+QG A++VYR+
Sbjct:      23 LLGLTVVAMTGAWLGMRYGGIAWE-SALQFNVHPLCMVIGLVFLQGDALLVYRV--FRNE 79

Query:      78 SKLLMKSIHAGLNAVAAILAIISVVAVFENHNVNNIANMYSLSHSWGVLIAVICVYLQLLS 137
              +K  K +H L+ A+++ ++VAVFE+H  A++YSLHSW G++  + Q L
Sbjct:      80 AKRTTKVHLGLHLHVAFVIALVGLVAVFEHHRKKGYADLYSLHSWCGILVFALFFAQWLV 139

Query:      138 GFSVFLLPWAPLSLRAFLMPIHVYSGIVIFGTVIATALMGLTEKLIFSLRDPAYSTFPPE 197
              GFS FL P A SLR+  P HV+ G IF  +ATAL+GL E L+F L  YSTF PE
Sbjct:      140 GFSFFLFGASFSLSRYSRPQHVFFGAAIFLLSVATALLGLKEALLFEL-GTKYSTFEPE 198

Query:      198 GVFVNTLGLLLILVFGALIFWIVTRPQWKRPKEPNSTIL 235
              GV N LGLL+  F ++ +I+TR WKRP +  L
Sbjct:      199 GVLANVLGLLLAATVTVVLYILTRADWKRPQAEQAL 236

```

Report for DKFZphfbr2_7e22.2

[illegible]

(No Pfam data available for DKFZphfbr2_7e22.2)

DKFZphfbr2_7j4

group: brain derived

DKFZphfbr2_7j4 encodes a novel 233 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit

Sequenced by GBF

Locus: unknown

Insert length: 1050 bp

Poly A stretch at pos. 1027, polyadenylation signal at pos. 1007

```

1 GGGGACACAA AGGGGTGGTC ACCCTGCCCT CACCTTGACC TGTAAGTTGC
51 CTAGGACAGT GGCCTGGTCC CAGGGGCTGT TGTGGGGAGT TGAAGAACAC
101 CCTGGCCTCC TCCATCATGT CGGCCAAGAG GGCAGAATTG AAGAAAACAC
151 ATCTGTGCAA GAACTACAAG GCAGTTTGCC TGGAAATTGA GCCAGAGCCG
201 ACCAAAACAT TTGATTACAA AGCAGTTAAA CAAGAAGGGC GGTTTACCAA
251 AGCAGGAGTG ACACAGGACC TAAAGAATGA ACTCAGGGAA GTGAGAGAAG
301 AGCTCAAGGA GAAATGGAG GAGATAAAAC AGATAAAGGA TCTAATGGAC
351 AAGGATTTTG ATAACTTCA CGAATTTGTG GAAATTATGA AGGAAATGCA
401 GAAAGATATG GATGAGAAGA TGGACATTTT AATAAATACA CAGAAGAACT
451 ATAAGCTTCC CCTTAGAAGA GCACCAAGG AGCAGCAGGA ACTCAGGCTG
501 ATGGGAAAGA CTCACAGAGA ACCACAGCTC AGGCCCAAGA AAATGGATGG
551 AGCCAGTGGA GTCATGGAG CACCTGTGTC TCTTCACAAG AAGACGATGG
601 CACCACAAA AAAAAACAG GGCTCACTGG ATCCCCTTCA TCACTGTGGG
651 ACCTGCTGCG AGAAATGTTT GTTGTGTGCT CTAAAGAACA ACTACAATCG
701 GGGGAACATT CCTTCAGAGG CCTCAGGCCT TTACAAAGGT GGAGAGGAGC
751 CAGTGACCAC CCAACCTTCT GTGGGCCACG CTGTGCCTGC CCCAAAGTCC
801 CAGACTGAGG GAAGGTGAAG CTTAAGTGCC AGCTTGAAAT GAGAGTAAAG
851 AAGATACAGA GCAACAGTG TTTCAGAAAC TGTCTGCCC TGGGTGTGAT
901 TCTTTGGCTT CAATTGAAG GAGGAGGAAT GATGGGATTT CATATTTTAT
951 TTCACACCAG TTCCTCCTTG TTTCATCTCT TTGCTAAGCT GGCTGCTTCT
1001 ACCATCTAAT AAATAATTGG CCAAGTTAAA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 117 bp to 815 bp; peptide length: 233
 Category: putative protein

```

1 MSAKRAELKK THLCKNYKAV CLELKPEPTK TFDYKAVKQE GRFTKAGVTQ
51 DLKNELEVR EELKEKMEEI KQIKDLMDKD FDKLHEFVEI MKEMQKDMDE
101 KMDILINTQK NYKLPLRRAP KEQQELRLMG KTHREPQLRP KKMDGASGVN
151 GAPCALHKKT MAPQKTKQGS LDPLHHCCTC CEKCLLCALK NNYNRGNIPS
201 EASGLYKGGE EPVTTQPSVG HAVPAPKSQT EGR

```

BLASTP hits

Entry JC2223 from database PIR:
 major surface glycoprotein 3 - Pneumocystis carinii (fragment)
 Score = 109, P = 3.5e-04, identities = 41/136, positives = 67/136

Alert BLASTP hits for DKFZphfbr2_7j4, frame 3

TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence., N = 1, Score = 109, P = 0.00024

>TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence.
Length = 196

HSPs:

Score = 109 (16.4 bits), Expect = 2.4e-04, P = 2.4e-04
Identities = 41/134 (30%), Positives = 67/134 (50%)

Query: 14 CKN-YKAVCLELKEPTKTFDYKAVKQEGRFKA-GVTQDLKNELREVREELKEKMEEIK 71
CK K C ELK + K VK+ TK G ++LK++++ E KE++E K
Sbjct: 22 CKTELKKYCEELKEADGLKVNDK-VKEICDDTKRDGKCKELKDKVKKELETFKEELE--K 78

Query: 72 QIKDLMKDFDKLHEFVEIMKEMQKMDKMDILINTQKNYKLPLRRAPKEQQELRLMGK 131
+KD+ D++ +K E +++E D D K + + YKL +R E LR +GK
Sbjct: 79 ALKDIDENCEKYEKCILLEETNHD-DVKKNCVKLREGCYKLKRKVA-EDLLLRALGK 136

Query: 132 THREPQLRPKKMDGAS 147
+ + K D S
Sbjct: 137 DVKNGECEKMKMDVCS 152

Pedant information for DKFZphfbr2_7j4, frame 3

Report for DKFZphfbr2_7j4.3

[LENGTH] 233
[MW] 26533.95
[pI] 9.18
[PROSITE] MYRISTYL 3
[PROSITE] CK2_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 3
[KW] All_Alpha
[KW] LOW_COMPLEXITY 14.59 %
[KW] COILED_COIL 13.73 %

SEQ MSAKRAELKKTHLCKNYKAVCLELKEPTKTFDYKAVKQEGRFKA-GVTQDLKNELREVR
SEGxx
PRD ccchhhhhhhhhhhccchhhhhhhcc
COILSCCCCCCCCCCCC

SEQ EELKEKMEEIKQIKDLMKDFDKLHEFVEIMKEMQKMDKMDILINTQKNYKLPLRRAP
SEGxx
PRD hhhhhhhhhhhhhhhhhccchhh
COILS CC

SEQ KEQQELRLMGKTHREPQLRPKKMDGASGVNGAPCALHKKTMAPQKTKQGSLDPLHHCCTC
SEG
PRD hhhhhhhhhhhhhhhhhcc
COILScc

SEQ CEKLLCALKNYNRGNIPSEASGLYKGGEPTVTPQSVGHAVPAPKSQTEGR
SEG
PRD chhhhhhhhhhhhhhhhhcc
COILScc

Prosites for DKFZphfbr2_7j4.3

PS00005	2->5	PKC_PHOSPHO_SITE	PDOC00005
PS00005	108->111	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_7j4.3)

DKFZphfbr2_82c20

group: transmembrane protein

DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with very weak similarity to C. elegans cosmid D1007.

The novel protein contains 7 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans D1007.5 ;
membrane regions: 7
Summary DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with
similarity to a hypothetical C.elegans protein.

similarity to C.elegans D1007.5

complete cDNA (Bp 1-100 GC ritch), complete cds,
potential start at Bp 128 matches Kozak consensus PyNNatgG,
EST hits, localisation? primer B of STS doesn't match perfect!
TRANSMEMBRANE 7

Sequenced by DKFZ

Locus: /map="109.9 cR from top of Chr1 linkage group"???

Insert length: 1804 bp

Poly A stretch at pos. 1794, no polyadenylation signal found

```

1 CGGCGGGAGC GCGCGGCTGA TACCCGGGAC TGGGCTGCGG CGGTTAGTCC
51 TCTCCCGGCC GCCGTCGCCT CCGACATATT GCTCGCAGGA GCTGCGGCGG
101 CGAAGCGGAG AGCACCAGGG GGAGGAGATG GGAGGACGAA GAGGTCCCAA
151 CAGGACATCT TACTGTCGAA ATCCGCTCTG TGAGCCGGGA TCCTCGGGGG
201 GCTCTAGTGG AAGCCACACT TCCAGTGCAT CGGTGACCAG TGTTCTGTTCC
251 CGCACCAGGA GCAGTTCTGG AACAGGCCTC TCCAGCCCTC CTCTGGCCAC
301 CCAAAGTGT GTGCCCTCTAC AGCACTGCAA GATCCCCGAG CTGCCAGTCC
351 AGGCCAGCAT TCTGTTTGAG TTGCAGCTCT TCTTCTGCCA GCTCATAGCA
401 CTCTTCGTCC ACTACATCAA CATCTACAAG ACAGTGTGGT GGTATCCACC
451 TTCCACACCA CCCTCCACCA CCTCCCTGAA CTTCATCTG ATCGACTTCA
501 ACTTGCTGAT GGTGACCACC ATCGTTCTGG GCGCCGCTT CATTTGGTCC
551 ATCGTGAAGG AGGCCTCTCA GAGGGGGAAG GTCTCCCTCT TTCGCTCCAT
601 CCTGCTGTTC CTCACCTCGT TCACCGTTCT CACGGCAACA GGCTGGAGTC
651 TGTGCCGATC CCTCATCCAC CTCTTCAGGA CCTACTCCTT CCTGAACCTC
701 CTGTTCTCT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCAGCTGAA
751 TTGCGACCTC CGCAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC
801 CCCGGGAGGC GGTGAGTGGC CTGGCAAAGA GCGGGGACTA CCTCCTGACA
851 CTGCGGGAGA CGTGAAGCA GCACACAAGA CAGCTGTATG GCCCGGACGC
901 CATGCCACCC CATGCCTGCT GCCTGTCAAC CAGCCTCATC CGCAGTGAGG
951 TTGAGTTTCT CAAGATGGAC TTCAACTGGC GCATGAAGGA AGTGCTCGTC
1001 AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCTG TCTGGTTCGT
1051 GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGAA CTCTTCCTGC
1101 TGGTGTCCAT CAGCACCTCC GTGATCCTCA TGCAGCACCT GCTGCCTGCC
1151 AGCTACTGTG ACCTGCTGCA CAAGGCCGCC GCCCATCTGG GCTGTTGGCA
1201 GAAGGTGGAC CCAGCGCTGT GCTCCAACGT GCTGCAGCAC CCGTGGACTG
1251 AAGAATGCAT GTGGCCGCAG GCGGTGCTGG TGAAGCACAG CAAGAAGCTC
1301 TACAAAGCCG TAGGCCACTA CAACGTGGCT ATCCCTCTG ACGTCTCCCA
1351 CTTCCGCTTC CATTTCTTTT TCAGCAAACC TCTGCGGATC CTCAACATCC
1401 TCCTGCTGCT GGAGGGCGCT GTCATTGTCT ATCAGCTGTA CTCCTTAATG
1451 TCCTCTGAAA AGTGGCACCA GACCATCTCG CTGGCCCTCA TCCTCTTCAG
1501 CAACTACTAT GCCTTCTTCA AGCTGCTCCG GGACCCGTTG GTATTGGGCA
1551 AGGCCTACTC ATACTCTGCT AGCCCCCAGA GAGACCTGGA CCACCGTTTC
1601 TCCTGAGCCC TGGGGTCACC TCAGGGACAG CGTCCAGGCT TCAGCCAAGG
1651 GCTCCCTGGC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA
1701 GACAAAAAAA TCCACCAGAG CTTGTATTTT TTGTTACGTA CTGTTTCTTT
1751 GATAATTGAT GTGATAAGGA AAAAAGTCCT ATTTTATATC TCCCAAAAAA
1801 AAAA

```

BLAST Results

Entry HS285343 from database EMBL:
human STS WI-17488.

Score = 1225, P = 1.3e-50, identities = 263/281

Medline entries

No Medline entry

Peptide information for frame 2

```

1 MGGRRGPNRT SYCRNPCEP GSSGGSSGSH TSSASVTSVR SRTSSSSGTG
51 LSSPPLATQT VVPLQHCKIP ELPVQASILF ELQLFFCQLI ALFVHYINIY
101 KTVWVYPPSH PPSHTSLNFH LIDFNLMVT TIVLGRRFIG SIVKEASQRG
151 KVSFLFRSILL FLTRFTVLTA TGWSLCRS LIHLFRTYSFLN LLFLCYPFGM
201 YIPFLQLNCD LRKTSLFNHM ASMGPREAVS GLAKSRDYLL TLRETWKQHT
251 RQLYGPDAMP THACCLSPSL IRSEVEFLKM DFNWRMKEVL VSSMLSAYYV
301 AFVPVWFVKN THYYDKRWSC ELFLVSIST SVILMQHLLP ASYCDLLHKA
351 AAHLGCWQKV DPALCSNVLQ HPWTEECMWP QGVLVKHSKN VYKAVGHYNV
401 AIPSDVSHFR FHEFFSKPLR ILNILLLEG AVIVYQLYSL MSSEKWHQTI
451 SLALILFSNY YAFFKLLRDR LVLGKAYSYS ASPQRDLDR FS

```

ORF from 128 bp to 1603 bp; peptide length: 492
 Category: similarity to unknown protein
 Prosite motifs: LEUCINE_ZIPPER (210-232)
 LEUCINE_ZIPPER (210-232)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82c20, frame 2

TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid
 D1007., N = 2, Score = 247, P = 4.6e-29

>TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007.
 Length = 512

HSPs:

Score = 247 (37.1 bits), Expect = 4.6e-29, Sum P(2) = 4.6e-29
 Identities = 58/204 (28%), Positives = 102/204 (50%)

```

Query: 291 VSSMLSAYYVAFVPVWFVKNTHYYDKRWSCFLVLSISTSVILMQHLLPASVCDLLHKA 350
      +S ML +V F + ++ W C+L ++V ++ + + +L P +Y DLLH+A
Sbjct: 299 LSIMLPCIFVPFKTSQGIPQKILINEVWECQLAIVVGLTAFSLYVAYLSPLNYDLLHRA 358

Query: 351 AAHLGCWQKVDPAL----CSNVLQHPWTEECMWPQGVLVKHSKN-VYKAVGHYNV---- 400
      A HLG W +++ P + + PW+E C++ G V+ Y+A ++
Sbjct: 359 AIHLGSWHQIEGPRIGHTGSMSSAPTWPSEFCLYNDGETVQMPDGRACYRAKSSNSIRTVA 418

Query: 401 AIPSDVSHFRFHEFFSKPLRILNILLLEGAVIVYQLYSLMSSEKWHQTI SLALILFSNY 460
      A P H F KP ++NI+ E +I Q + L+ + W ++ L++F+NY
Sbjct: 419 AHPESSRHNTFFKVLKPNLNINIMCSFEFLIFIQFWMLVLTNDWQHIVTFVLLMFANY 478

Query: 461 YAFFKLLRDRDLVLGKAYSYSASPQRDL 487
      F KL +D+++L + Y S Q DL
Sbjct: 479 LLFAKLFDKILSRIEPS---QEDL 502

```

Score = 178 (26.7 bits), Expect = 4.3e-21, Sum P(2) = 4.3e-21
 Identities = 50/179 (27%), Positives = 90/179 (50%)

```

Query: 262 HACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYVAFVPVWFV--KNTHYYDKR-- 317
      H C SP+ IR E++ L D R+K+ + + + +A+ +P EV K + ++
Sbjct: 262 HMCSDSPAQIREEQVLIDDLVLRVKKSFAGVSTAFSLIMLPCIFVPFKTSQGIPQKIL 321

Query: 318 ----WSCFLVLSISTSVILMQHLLPASVCDLLHKAHAAHLGCWQKVDPAL----CSNV 368
      W C+L ++V ++ + + +L P +Y DLLH+AA HLG W +++ P + +
Sbjct: 322 INEVWECQLAIVVGLTAFSLYVAYLSPLNYDLLHRAAIHLGSWHQIEGPRIGHTGSMSS 381

Query: 369 LQHPWTEECMWPQGVLVKHSKN-VYKAVGHYNV-AIPSDVSHFRFHEFFSKPLRILNILL 426
      PW+E C++ G V+ Y+A ++ + + R + FF K LR N L+
Sbjct: 382 APTPWSEFCLYNDGETVQMPDGRACYRAKSSNSIRTVAHPESSRHNTFF-KVLKPNLI 440

```


SEQ ASPQRDLDRFS
 SEG
 PRD ccchhhhhccc
 MEM

Prosites for DKFZphfbr2_82c20.2

PS00001	8->12	ASN_GLYCOSYLATION	PDOC00001
PS00002	47->51	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	212->216	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	316->320	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	443->446	PKC_PHOSPHO_SITE	PDOC00005
PS00006	241->245	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	28->34	MYRISTYL	PDOC00008
PS00008	48->54	MYRISTYL	PDOC00008
PS00008	231->237	MYRISTYL	PDOC00008
PS00009	2->6	AMIDATION	PDOC00009
PS00009	134->138	AMIDATION	PDOC00009
PS00029	168->190	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_82c20.2)

DKFZphfbr2_82e17

group: transmembrane protein

DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with very weak similarity to C. elegans cosmid R01B10.

The novel protein contains 6 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans "R01B10.5" ;
membrane regions: 6
Summary DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with
similarity to a hypothetical C.elegans protein.

similarity to C.elegans "R01B10.5"

complete cDNA, EST HS763158 extends the sequence, complete cds, EST
hits
six potential transmembrane domains

Sequenced by DKFZ

Locus: /map="779_C_?; 818_A_1; 877_C_1; 734_C_12; 760_E_11; 171.7 cR from top of Chr14 linkage
group"

Insert length: 1618 bp
Poly A stretch at pos. 1608, polyadenylation signal at pos. 1588

```

1 CTGATCTAGT GCTTCTCGAA AAAACCTTC AGGCGGCCCA TGGCTGTCGA
51 TATTCACCCA GCATGCCTTG GACTTTATTG TGGGAAGACC CTATTATTTA
101 AAAATGGCTC AACTGAAATA TATGGAGAAT GTGGGGTATG CCCAAGAGGA
151 CAGAGAACGA ATGCACAGAA ATATTGTCAG CCTTGACAG AATCTCCTGA
201 ACTTTATGAT TGGCTCTATC TTGGATTAT GGCAATGCTT CCTCTGGTTT
251 TACATTGGTT CTTCATTGAA TGGTACTCGG GGAAAAGAG TTCCAGCGCA
301 CTTTTCACAC ACATCACTGC ATTATTGAA TGCAGCATGG CAGCTATTAT
351 CACCTTACTT GTGAGTGATC CAGTTGGTGT TCTTTATATT CGTTCATGTC
401 GAGTATTGAT GCTTCTGAC TGGTACACGA TGCTTTACAA CCCAAGTCCA
451 GATTACGTTA CCACAGTACA CTGTACTCAT CAAGCCGCTT ACCCACTATA
501 TACCATTGTA TTTATCTATT ACGCATTTCTG CTGGGTATTA ATGATGCTGC
551 TCCGACCTCT TCTGGTGAAG AAGATTGCAT GTGGGTAGG GAAATCTGAT
601 CGATTTAAAA GTATTTATGC TGCACTTTAC TTCTTCCCAA TTTTAACCGT
651 GCTTCAGGCA GTTGGTGGAG GCCTTTTATA TTACGCCTTC CCATACATTA
701 TATTAGTGTT ATCTTTGGTT ACTCTGGCTG TGTACATGTC TGCTTCTGAA
751 ATAGAGAACT GCTATGATCT TCTGGTCAGA AAGAAAAGAC TTATTGTTCT
801 CTTACAGCCAC TGGTTACTTC ATGCCTATGG AATAATCTCC ATTTCCAGAG
851 TGGATAAACT TGAGCAAGAT TTGCCCTTTT TGGCTTTGGT ACCTACACCA
901 GCCCTTTTTC ACTTGTTCAC TGCAAAATTT ACCGAACCTT CAAGGATACT
951 CTCAGAAGGA GCCAATGGAC ACTGAGTGTA GACATGTGAA ATGCCAAAAA
1001 CCTGAGAAGT GCTCCTAATA AAAAAGTAAA TCAATCTTAA CAGTGTATGA
1051 GAACTATTCT ATCATATATG GGAACAAGAT TGTCAAGTATA TCTTAATGTT
1101 TGGGTTTGTG TTTGTTTGTG TTATGGTTAG ACTTACAGAC TTGGAAAATG
1151 CAAAACCTCTG TAATACTCTG TTACACAGGG TAATATTATC TGCTACACTG
1201 GAAGGCCGCT AGGAAGCCCT TGCTTCTCTC AACAGTTCAG CTGTTCTTTA
1251 GGGCAAAATC ATGTTTCTGT GTACCTAGCA ATGTGTTCCC ATTTTATTAA
1301 GAAAAGCTTT AACACGTGTA ATCTGCAGTC CTTAACAGTG GCGTAATTGT
1351 ACGTACCTGT TGTGTTTCAG TTTGTTTTTC ACCTATAATG AATTGTAAAA
1401 ACAAACATAC TTGTGGGGTC TGATAGCAAA CATAGAAATG ATGTATATTG
1451 TTTTTTGTTA TCTATTTATT TTCATCAATA CAGTATTTTG ATGTATTGCA
1501 AAAATAGATA ATAATTTATA TAACAGGTTT TCTGTTTATA GATTGGTTCA
1551 AGATTGTTT GGATTATTGT TCCTGTAAAG AAAACAATAA TAAAAAGCTT
1601 ACCTACATAA AAAAAAAA

```

BLAST Results

Entry HS981146 from database EMBL:
human STS WI-6253.
Length = 208
Minus Strand HSPs:
Score = 1040 (156.0 bits), Expect = 1.9e-40, P = 1.9e-40

Identities = 208/208 (100%), Positives = 208/208 (100%), Strand = Minus
/ Plus

Entry HSG20716 from database EMBL:

human STS A006D06.

Length = 195

Minus Strand HSPs:

Score = 975 (146.3 bits), Expect = 1.8e-37, P = 1.8e-37

Identities = 195/195 (100%), Positives = 195/195 (100%), Strand = Minus
/ Plus

Medline entries

No Medline entry

Peptide information for frame 1

```

1 MAVDIQPAQL GLYCGKTLLE KNGSTEIYGE CGVCPRGQRT NAQKYCQPCT
51 ESPELYDWLY LGFMAMLPV LHWFFIEWYS GKSSSSALFQ HITALFECSM
101 AAIITLLVSD PVGVLYIRSC RVLMLSDWYT MLYNPSPDYV TVHCTHEAV
151 YPLYTIVFIY YAFCLVLMML LRPLLVKKIA CGLGKSDRFK SIYAALYFFP
201 ILTVLQAVGG GLLYYAFPIY ILVLSLVTLA VYMSASEIEN CYDLLVRKKR
251 LIVLFSHWLL HAYGIISISR VDKLEQDLPL LALVPTPALF YLFTAKFTEP
301 SRILSEGANG H

```

ORF from 40 bp to 972 bp; peptide length: 311
Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82e17, frame 1

TREMBL:AF068718.5 gene: "R01B10.5"; Caenorhabditis elegans cosmid
R01B10., N = 1, Score = 399, P = 1.4e-36

>TREMBL:AF068718.5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10.
Length = 670

HSPs:

Score = 399 (59.9 bits), Expect = 1.4e-36, P = 1.4e-36
Identities = 95/280 (33%), Positives = 152/280 (54%)

```

Query: 2 AVDIQPAQLGLYCGKTLLEKFN-----GSTIYGE CGVCPRGQRTNAQKYCQPC 49
      A IQP+CLG +CG+T+L N GST + CG C G R NA C+ C
Sbjct: 292 ASTIQPSCLG-FCGRTVLVGNYSDEATTAAAGSTSL-SRCGPCSFGYRNAMSI CESC 349

Query: 50 TESPELYDWLYLGFMAMLPVLHWWFFIEWYSGKSSSALFQ---HITALFECSMAAIITL 106
      + YDW+YL F+A+LPL+LH FI + K + ++ ++ + E +A +I +
Sbjct: 350 DTPLPQYDWMYLLFIALLPLLHMQFIR-IARKYCRTRYEVSEYLCVILENVIACVIAV 408

Query: 107 LVSDPVGVLVYIRSCRVLMLSDWYTMLYNPSPDYVTTVHCTHEAVYPLYTIVFIYYAFCLV 166
      L+ P ++ C + +WY YNP Y T+ CT+E V+PLY+I FI++ +
Sbjct: 409 LIYPPRFTFFLNGCSKTDIKEWYPACYNPRIGYTKTMRCTYEVVFPYLSITFIHHLILIG 468

Query: 167 LMMLLRPLLVKKIAACGLGKSDRFKSIYAALYFFPIILTVLQAVGGGLLYYAFPIIILVLSL 226
      +++LR L + L K+ K YAA+ PIL V+ AV G+++Y FPYI+L+ SL
Sbjct: 469 SILVLRSTLYCVL---LYKTYNGKPFYAAIVSVPILAVIHAVLSGVVVFYTFPYILLIGSL 525

Query: 227 VTLAVYMSASEIENCYDLLVR---KKRLIVLFSHWLLHAYGIISI 268
      + +++ +++VR LI L L+ ++G+I+I
Sbjct: 526 WAMCFHLALEGKRPLKEMIVRIATSPTHLIFLSITMLMLSFGVIAI 571

```

Pedant information for DKFZphfbr2_82e17, frame 1

Report for DKFZphfbr2_82e17.1


```

[LENGTH]      311
[MW]           35239.14
[pI]           7.91
[HOMOL]        TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10. 9e-36

[PROSITE]      AMIDATION      1
[PROSITE]      MYRISTYL       3
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 3
[PROSITE]      PKC_PHOSPHO_SITE 4
[PROSITE]      ASN_GLYCOSYLATION 1
[KW]           TRANSMEMBRANE 6
[KW]           LOW_COMPLEXITY 7.72 %

SEQ      MAVDIQPACLGLYCGKTLLEFKNGSTEIYGECGVCPRGQRTNAQKYCQPCTESPELYDWLY
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhh
MEM      .....MMMMMM

SEQ      LGFMAMLPVLHWHFFIEWYSGKKSSSALFQHITALFECSMAAIITLLVSDPVGVLIRSC
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeeeece
MEM      MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      RVLMLSDWYTMLYNPSPDYVTTVHCTHEAVYPLYTIVFIYYAFCLVLMMLRPLLVKKIA
SEG      .....xxxxxxxxxxxxx...
PRD      eeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhe
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      CGLGKSDRFKSIYAALYFFPILTVLQAVGGGLLYYAFPIIILVLSLVTLAVYMSASEIEN
SEG      .....
PRD      eccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      CYDLLVRKRLIVLFHSHWLLHAYGIISISRVDKLEQDPLLLALVPTPALFYLFYFTAKFTEP
SEG      .....xxxxxxxxxxxxx...
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      SRILSEGANGH
SEG      .....
PRD      ceeeecccccc
MEM      MM.....

```

Prosite for DKFZphfbr2_82e17.1

PS00001	22->26	ASN_GLYCOSYLATION	PDOC00001
PS00004	82->86	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	119->122	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	294->297	PKC_PHOSPHO_SITE	PDOC00005
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	269->273	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00009	80->84	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82e17.1)

DKFZphfbr2_82e4

group: signal transduction

DKFZphfbr2_82e4 encodes a novel 473 amino acid protein with strong similarity to the calmodulin-binding proteins.

The novel protein is similar to human and rat Ca²⁺/calmodulin-dependent protein kinase (EC 2.7.1.123), rat calmodulin-binding protein, calmodulin binding protein kinase of Fugu rupies and Rattus norvegicus calcium/calmodulin-dependent protein kinase I. Calmodulin is the archetype of the family of calcium-modulated proteins of which nearly 20 members have been found. Calmodulin is involved in regulation of growth and cell cycle as well as in signal transduction and the synthesis and release of neurotransmitters. The novel protein seems to be involved in calmodulin-mediated pathways in human neuronal cells.

The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

strong similarity to calmodulin-binding proteins

complete cDNA, complete cds, EST hits
splice variant in comparison to rat I56542
ESTs HS2Z54543/HS1141907 define splice variant
see also DKFZphfbr2_82g20 unspliced form

Sequenced by DKFZ

Locus: /map="200.5 cR from top of Chr3 linkage group"

Insert length: 2923 bp

Poly A stretch at pos. 2913, polyadenylation signal at pos. 2890

```
1 ATGCTGGAGG TTCGCTAGCC GAAGCGGCTG CATCTGGCGC CGCGTCTGCC
51 CCGCGTGCTC GGAGCGGATT CTGCCC GCCG TCCCGGAGC CCTCGCGGCC
101 CCGCTGAGCC CGCGATCACT TCCTCCCTGT GACCAACCGG CGCTGCAGGT
151 TAGAGCCTGG CAATGCCGTT TGGGTGTGTG ACTCTGGGTG ACAAGAAGAA
201 CTATAACCAG CCATCGGAGG TGA CTGACAG ATATGATTTG GGACAGGTCA
251 TCAAGACTGA GGAGTTTTGT GAAATCTTCC GGGCCAAGGA CAAGACGACA
301 GGCAAGCTGC ACACCTGCAA GAAGTTCCAG AAGCGGGACG GCCGCAAGGT
351 GCGGAAAGCT GCCAAGAAGC AGATAGGCAT CCTCAAGATG GTGAAGCATC
401 CCAACATCCT ACAGCTGGTG GATGTGTTTG TGACCCGCAA GGAGTACTTT
451 ATCTTCCTGG AGCTGGCCAC GGGGAGGGAG GTGTTTGACT GGATCTGGA
501 CCAGGGCTAC TACTCGGAGC GAGACACAAG CAACGTGGTA CGGCAAGTCC
551 TGGAGGCCGT GGCCTATTTG CACTCACTCA AGATCGTGCA CAGGAATCTC
601 AAGCTGGAGA ACCTGGTTTA CTACAACCGG CTGAAGAAGT CGAAGATTGT
651 CATCAGTGAC TTCCATCTGG CTAAGCTAGA AAATGGCCTC ATCAAGGAGG
701 CCTGTGGGAC CCCCAGATAT CTGGGCAACC CACCTTTCTA TGAGGAGGTG
751 GAAGAAGATG ATTATGAGAA CCATGATAAG AATCTCTTCC GCAAGATCCT
801 GGCTGGTGAC TATGAGTTTG ACTCTCCATA TTGGGATGAT ATTTCCGAGG
851 CAGCCAAAGA CCTGGTCACA AGGCTGATGG AGGTGGAGCA AGACCAAGCG
901 ATCACTGCAG AAGAGGCCAT CTCCCATGAG TGGATTCTTG GCAATGTCTG
951 TTCTGATAAG AACATCAAGG ATGGTGCTTG TGCCAGATT GAAAAGAACT
1001 TTGCCAGGGC CAAGTGGAAG AAGGCTGTCC GAGTGACCAC CCTCATGAAA
1051 CGGCTCCGGG CACCAGAGCA GTCCAGCACG GCTGCAGCCC AGTCGGCCTC
1101 AGCCACAGAC ACTGCCACCC CCGGGGCTGC AGGTGGGGCC ACAGCTGCAG
1151 CTGCGAGTGG AGCTACCTCA GCCCTGAGG GTGATGTCTG TCGTGCTGCA
1201 AAGAGTGATA ATGTGGCCCC CGCAGACCGT AGTGCCACCC CAGCCACAGA
1251 TGGAAGTGCC ACCCCAGCCA CTGATGGCAG TGTACCCCA GCCACCGATG
1301 GAAGCATCAC TCCAGCCACT GATGGGAGTG TCACCCAGC CACTGACAGG
1351 AGCGCTACTC CAGCCACTGA TGGGAGAGCC ACACCAGCCA CAGAAGAGAG
1401 CACTGTGCC ACCACCCAAA GCAGTGCCAT GCTGGCCACC AAGGCAGCTG
1451 CACCCCTGTA GCCGGCTATG GCCAGCCGG ACAGCACAGC CCCAGAGGGC
1501 GCCACAGGCC AGGCTCCACC CTCTAGTAAA GGGGAAGAGG CTGCTGGTTA
1551 TGCCCAAGAG TCTCAAAGGG AGGAGGCCAG CTGAGTAGGC AGCCTGGTGA
1601 GGGGGGGCAG GGGATGGGCA GGAGGGTGGG AGAGTGGATG AGGGGCTTCT
1651 CACTGTACAT AGAGTCACTG GCATGATGCC CTCGCTCCCC CATGCCCCCA
1701 CATCCCAAGT GGGCATAACT AGGGGTACAG GGAGAGCAGT CTCGTCTCCT
1751 GTGTGTATGT GTGTGAGTGG TGGGCAGGCC AGTGGCAGGG CCGGCCCCAG
1801 CCCCTGCATG GATTCTTTGT GGCTTTTCTG TCTTTTGCTA GCTTCACCAT
1851 TTTCTGTCTC TTGTGGGATG CTGCTCTAGG GATACTCAGG GGGCTCCTGC
1901 TCTCCTTCCC CTTCCTTCT TGCCTCACC A TTCCCTAGG CAGGCCCTGC
1951 AGGTCCACCA CTCTCCAGG CCTTAACTT GGGCGGCCTT GCCCTGAGAG
2001 CTGGTCTCTC AGCGAGGCC TGTCAGCGGT CTTAGGCTCC TGCACATGAA
2051 GGTGTGTGCC TGTGGTGTGT GGGCTGTCTT AGGAGCAGAT ACAGGCTGGT
2101 ATAGAGGATG CAGAAAGGTA GGGCAGTATG TTTAAGTCCA GACTTGGCAC
2151 ATGGCTAGGG ATACTGCTCA CTAGCTGTGG AGGTCCTCAG GAGTGGAGAG
2201 AATGAGTAGG AGGGCAGAAG CTTCCATTTT TGTCTTCTCT AAGACCTGT
```



```

2251 TATTGTGTT ATTTCTGCC TTTCCGAGTC CTGCAGTGGG CTGCCCTGTA
2301 CCCTGAACCT CATGAGCCTC TAAGGGAAAG GAGGAACAAT TAGGACGTGG
2351 CAATGAGACC TGGCAGGGCA GAGTACAAGC CCAGCACCCA GTGTCCCAGC
2401 CTTACTGGGT CCTTACCCTG GGCCAAACAG GGAGGGCTGA TACCTCCTTG
2451 CTCTTCCTAG ATGCCACCT CCTACAATCT CAGCCCACAA GTCCTCTCCA
2501 CCCTAGGGGG CTTGCTGCAT GGCAATAACT CATAATCTGA TTTGGAGGTT
2551 TGCCCTTTAC AGGGGCAGAT TTTCTGCTCA GTTCAACAAT GAAATGAAGA
2601 GGAACCTCCT CTTTCTACAG CTCACTTCTA TCAGAGGGCC AGGTGCCTCA
2651 GAGCCACATT GAGTTGCTTT TTCTGGGATG AGGAAGTAGG GTTAAACTCC
2701 CCAGTTTCCT GAGGGAGGCT CCTGACAGGT GCCCTTTGTC AGACCCTACC
2751 ACAGCCTGGA TAGGCAGCCA CATTGGTCTT CGCCCTTGCT CGGCACTCCG
2801 TGGTGGTCTT GCCCTTCTCC CTGCATGCCT GTGGGTCTGC TCTGGTGTGT
2851 GAAGGTCGGT GGGTTAACTG TGTGCCTACT GAACCTGGCA AATAAACATC
2901 ACCCTGCAAA GCCAAAAAA AAA

```

BLAST Results

Entry HS452352 from database EMBL:

human STS WI-15318.

Length = 350

Minus Strand HSPs:

Score = 1547 (232.1 bits), Expect = 5.2e-63, P = 5.2e-63

Identities = 331/348 (95%), Positives = 331/348 (95%), Strand = Minus /

P1

Medline entries

94110847:

J Neurosci 1994 Jan;14(1):1-13

IG5: a calmodulin-binding, vesicle-associated, protein kinase-like protein enriched in forebrain neurites.

Godbout M, Erlander MG, Hasel KW, Danielson PE, Wong KK, Battenberg EL, Foye PE,

Bloom FE, Sutcliffe JG

Peptide information for frame 1

```

1 MPFGCVTLGD KKNYNQPSV TDRYDLGQVI KTEEFCEIFR AKDKTTGKLIH
51 TCKKFQKRDG RKRKAANE IGILKMKVHP NILQLVDV FV TRKEYFIFLE
101 LATGREVFDFW ILDQGYYSER DTSNVVRQVL EAVAYLHSLK IVHRNLKLEN
151 LVYVNRKLSN KIVISDFHLA KLENGLIKEP CGTPEYLGNP PFYEEVEEDD
201 YENHDKNLFR KILAGDYEFD SPYWDDISQA AKDLVTRLME VEQDQRITAE
251 EAISHEWISG NAASDKNIKD GVCAQIEKNF ARAKWKKA VR VTTLMKRLRA
301 PEQSSTAAAQ SASATDTATP GAAGGATAAA ASGATSAPEG DAARAAKSDN
351 VAPADRSATP ATDGSATPAT DGSVTPATDG SITPATDGSV TPATDRSATP
401 ATDGRATPAT EESTVPTTQS SAMLATKAAA TPEPAMAQPD STAPEGATGQ
451 APPSSKGEEA AGYAQESQRE EAS

```

ORF from 163 bp to 1581 bp; peptide length: 473

Category: strong similarity to known protein

BLASTP hits

Entry S50193 from database PIR:

Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - rat

Length = 374

Score = 371 (130.6 bits), Expect = 2.2e-66, Sum P(2) = 2.2e-66

Identities = 74/176 (42%), Positives = 115/176 (65%)

Entry S57347 from database PIR:

Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - human

Length = 370

Score = 369 (129.9 bits), Expect = 4.6e-66, Sum P(2) = 4.6e-66

Identities = 74/176 (42%), Positives = 114/176 (64%)

Alert BLASTP hits for DKFZphfbr2_82e4, frame 1

PIR:I56542 calmodulin-binding protein - rat, N = 2, Score = 1246, P = 4e-228

TREMBLNEW:FRU010348.3 product: "calmodulin binding protein kinase";
Fugu rubripes UBE1-like gene, PRGFR2 gene and gene encoding calmodulin
binding protein kinase, clone 168J21, N = 2, Score = 846, P = 2.6e-139

TREMBL:RNPRKI_1 product: "protein kinase I"; Rattus norvegicus
calcium/calmodulin-dependent protein kinase I mRNA, complete cds., N =
2, Score = 364, P = 5.1e-63

>PIR:I56542 calmodulin-binding protein - rat
Length = 504

HSPs:

Score = 1246 (186.9 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 255/289 (88%), Positives = 259/289 (89%)

Query: 188 GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSPYWDNISQAADLVTRLMEVEQDQRI 247
GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSPYWDNISQAADLVTRLMEVEQDQRI
Sbjct: 216 GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSPYWDNISQAADLVTRLMEVEQDQRI 275

Query: 248 TAEAAISHEWISGNAASDKNIKDGVCQAQIEKNFARAKWKKAVRVTTLMKRLRAPEQSSTA 307
TAEAAISHEWISGNAASDKNIKDGVCQAQIEKNFARAKWKKAVRVTTLMKRLRAPEQS TA
Sbjct: 276 TAEAAISHEWISGNAASDKNIKDGVCQAQIEKNFARAKWKKAVRVTTLMKRLRAPEQSGTA 335

Query: 308 AAQSASATDGTATPGAAGGATAAASGATSAP-----GDAARAASDNVAPADRSAT 359
A +D ATPGAAGGA AAAA GA A GDA AAKSD++A ADRSAT
Sbjct: 336 AT-----SDAATPGAAGGAVAAAAGGAAPASGASATVGTGGDAGCAAKSDMASADRSAT 390

Query: 360 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQ 419
PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVP Q
Sbjct: 391 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPAQA 450

Query: 420 SSAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS 473
SSA A KAAATPEPA+AQPDSTA EGATGQAPPSSKGEEA G AQESQR E S
Sbjct: 451 SSAAPAAKAAATPEPAVAQPDSTALEGATGQAPPSSKGEEATGCAQESQRVETS 504

Score = 978 (146.7 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 186/187 (99%), Positives = 187/187 (100%)

Query: 1 MPFGCVTLGDKKNYNQPSVETDRYDLGQVVIKTEEFCEIFRAKDKTGKLTCKKFQKRDG 60
MPFGCVTLGDKKNYNQPSVETDRYDLGQV+KTEEFCEIFRAKDKTGKLTCKKFQKRDG
Sbjct: 1 MPFGCVTLGDKKNYNQPSVETDRYDLGQVVKTEEFCEIFRAKDKTGKLTCKKFQKRDG 60

Query: 61 RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER 120
RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER
Sbjct: 61 RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER 120

Query: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180
DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
Sbjct: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180

Query: 181 CGTPEYL 187
CGTPEYL
Sbjct: 181 CGTPEYL 187

Pedant information for DKFZphfbr2_82e4, frame 1

Report for DKFZphfbr2_82e4.1

[LENGTH] 473
[MW] 51208.89
[pI] 5.30
[HOMOL] PIR:I56542 calmodulin-binding protein - rat 0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YDL101c] 8e-26
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 5e-24
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 7e-23
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
7e-23
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 1e-21
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 1e-21

[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-16
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 3e-16
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 1e-15
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 3e-14
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YCR073c] 6e-11
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 8e-11
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YJL095w] 2e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 5e-07
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 8e-07
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 1e-05
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR523c] 2e-04
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-04
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[BLOCKS] BL00939F
[SCOP] dlgo1_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-62
[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 5e-59
[SCOP] dlkoa_ 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 1e-75
[SCOP] dlkoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 1e-72
[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 4e-65
[SCOP] dlirk_ 5.1.1.2.4 insulin receptor (Human (Homo sapiens)] 2e-56
[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mus)] 4e-71
[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Homo)] 1e-50
[SCOP] dlydre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bos)] 3e-70
[SCOP] dlfmk_ 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Homo)] 5e-49
[SCOP] dlcdkb_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Sus)] 2e-72
[SCOP] dlhcka3_ 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [human (Homo)] 5e-46
[SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 9e-42
[SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 1e-56
[SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 9e-52
[EC] 2.7.1.38 Phosphorylase kinase 3e-29
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 8e-66
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 2e-17
[EC] 2.7.1.117 Myosin-light-chain kinase 2e-38
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase (NADPH)] kinase 2e-17
[EC] 2.7.1.37 Protein kinase 6e-28
[PIRKW] phosphotransferase 8e-66
[PIRKW] nucleus 2e-24
[PIRKW] transferase 8e-30
[PIRKW] calcium 2e-27
[PIRKW] duplication 4e-19
[PIRKW] tandem repeat 2e-31
[PIRKW] phorbol ester binding 1e-16
[PIRKW] zinc 1e-16
[PIRKW] cell cycle control 2e-20
[PIRKW] serine/threonine-specific protein kinase 8e-66
[PIRKW] phospholipid binding 1e-16
[PIRKW] autophosphorylation 8e-66
[PIRKW] brain 1e-14
[PIRKW] heterotetramer 2e-16
[PIRKW] polymer 3e-29
[PIRKW] mitosis 2e-20
[PIRKW] magnesium 7e-22
[PIRKW] ATP 8e-66
[PIRKW] alternative initiators 1e-29

[PIRKW] phosphoprotein 8e-66
 [PIRKW] apoptosis 2e-31
 [PIRKW] glycoprotein 4e-19
 [PIRKW] skeletal muscle 3e-28
 [PIRKW] protein kinase 2e-28
 [PIRKW] testis 3e-28
 [PIRKW] signal transduction 1e-21
 [PIRKW] cAMP binding 1e-16
 [PIRKW] purine nucleotide binding 5e-25
 [PIRKW] structural protein 4e-19
 [PIRKW] calcium binding 3e-45
 [PIRKW] alternative splicing 3e-45
 [PIRKW] P-loop 5e-25
 [PIRKW] lipoprotein 2e-16
 [PIRKW] cardiac muscle 4e-19
 [PIRKW] muscle 3e-28
 [PIRKW] myristylation 2e-16
 [PIRKW] EF hand 5e-29
 [PIRKW] cell division 2e-38
 [PIRKW] calmodulin binding 8e-66
 [PIRKW] smooth muscle 7e-31
 [SUPFAM] fibronectin type III repeat homology 7e-31
 [SUPFAM] immunoglobulin homology 7e-31
 [SUPFAM] ribosomal protein S6 kinase II 3e-26
 [SUPFAM] calcium-dependent protein kinase 5e-29
 [SUPFAM] AMP-activated protein kinase 7e-22
 [SUPFAM] protein kinase akt 1e-14
 [SUPFAM] protein kinase SPK1 3e-20
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-36
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-45
 [SUPFAM] calmodulin repeat homology 5e-29
 [SUPFAM] protein kinase DUN1 2e-24
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-14
 [SUPFAM] death-associated protein kinase 2e-31
 [SUPFAM] myosin-light-chain kinase, nonmuscle 1e-29
 [SUPFAM] pleckstrin repeat homology 1e-14
 [SUPFAM] ankyrin repeat homology 2e-31
 [SUPFAM] protein kinase homology 8e-66
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-36
 [SUPFAM] twitchin 1e-18
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-16
 [SUPFAM] titin 4e-19
 [SUPFAM] protein kinase cdrl 2e-20
 [SUPFAM] kinase-related transforming protein 2e-38
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-66
 [SUPFAM] kinase interaction domain homology 2e-24
 [SUPFAM] protein kinase C mu 1e-16
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 3
 [PROSITE] CK2_PHOSPHO_SITE 10
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 11
 [PFAM] Eukaryotic protein kinase domain
 [KW] All_Alpha
 [KW] 3D
 [KW] LOW_COMPLEXITY 7.40 %

SEQ MPFGCVTLGDKKNYNQPSVETDRYDLGQVIKTEEFCEIFRAKDKTGKLTCKKFQKRDG
 SEG
 1a06-CEETTTGGGCCEEEECBCGGGGGEEEEETTTTCEEEEEEEEC---

 SEQ RKVRKAAKNEIGILKMVKHPNQLVDVFTVTRKEYFIFLELATGREVFDWILDQGYYSER
 SEG
 1a06- -----HHHHHHHHHCCTTTTBCCEEEEEETTEEEEECCCCCEHHHHHHHHTTTTBHH

 SEQ DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
 SEG
 1a06- HHHHHHHHHHHHHHHHHHHHCCCTTTTTTTTTEEECCCTTTTCEEECCCTTTTCHHHHHHCCC

 SEQ CGTPEYLGNNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSFYWDDISQAADLVTRLME
 SEG
 1a06- HHHHHHHHCCTTTT-----THHHHHHHHCCCCCTTTTTCCHHHHHHHHHHCT

 SEQ VEQDQRITAEAEISHEWISGNAASDKNIKDGVCQAQIEKNFARAKWKKAVRVTTLMKRLRA
 SEG
 1a06- TTGGGCCCHHHHHHTTTTTCCTCCCBHHHHHHHHHHHHHCCTTTTBTBHHHHHHHC..

 SEQ PEQSSTAAQASATDTATPGAAGGATAAASGATSAPEGDAARAASDNVAPADRSATP
 SEG
 1a06-

SEQ ATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQS
 SEG
 1a06-

SEQ SAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS
 SEG
 1a06-

Prosite for DKFZphfbr2_82e4.1

PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	91->94	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	118->121	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	264->267	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	454->457	PKC_PHOSPHO_SITE	PDOC00005
PS00005	467->470	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	118->122	CK2_PHOSPHO_SITE	PDOC00006
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	336->340	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00007	456->464	TYR_PHOSPHO_SITE	PDOC00007
PS00007	127->136	TYR_PHOSPHO_SITE	PDOC00007
PS00008	260->266	MYRISTYL	PDOC00008
PS00008	321->327	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00009	59->63	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_82e4.1

HMM_NAME	Eukaryotic protein kinase domain		
HMM	*YeigRiIGeGsFGtVYkCiWr.TGeIVAiKIIkkrsms.....FlREIq		
Query	24	YDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDGRKVRKAANEIG	72
HMM	IMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEwe		
Query	73	ILKMVKHPNQLQVDVfV-TRKEYFIFLELATGREVFDWILDQGYYSERD	121
HMM	IrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeN...gqIKICDFGLAR		
Query	122	TSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAK	171
HMM	qMnnYerMttfCGTPWY*		
Query	172	LEN--GLIKEPCGTPEY	186
HMM	*GepPFYd.....dnMemImrIiqrfripfWpnCSeElyDFMr		
Query	188	GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAADLVT	236
HMM	wCWnyDPekRPTFrQILnHPWF*		
Query	237	RLMEVEQDQRITAEEAISHEWI	258

DKFZphfbr2_82g14

group: transmembrane protein

DKFZphfbr2_82g14 encodes a novel 208 amino acid proline-rich protein without similarity to known proteins.

The protein contains one transmembrane domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

unknown prolin rich protein

membrane regions: 1

Summary DKFZphfbr2_82g14 encodes a novel 208 amino acid protein.

unknown prolin rich protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 2059 bp

Poly A stretch at pos. 2049, polyadenylation signal at pos. 2024

```
1 AGAAGTGGCA CTGCCAGCTG CCGAGGCGTT CGGTCTGCT GTTGCGGCCG
51 CTGCCCCAGG GCTGCGGGGA CGCTCCCGGA GCCCTGCCTG TCCCCTGTCC
101 ATCCAGGCCA GCAGCTGAAG GAGCCTCACC TGCCTCCCTT CTCTGAGTAG
151 CACGGATTGT AGGAGAAGCA GCGAAGATGT CCAGCGAGCC TCCCCCTCCT
201 TATCTTGGGG GCCCCACAGC CCCACTTCTG GAAGAGAAAA GTGGAGCCCC
251 GCCCCACCCA GGCCGTTTCT CCCCAGCTGT GATGCAGCCC CCTCCAGGCA
301 TGCCACTGCC CCCTGCGGAC ATTGCCCCC CACCCTATGA GCCGCCGGGT
351 CACCCAATGC CCCAGCCTGG CTTCATCCCA CCACACATGA GTGCAGATGG
401 CACCTACATG CCTCCGGGTT TCTACCCTCC TCCAGGCCCC CACCCACCCA
451 TGGGCTACTA CCCCCAGGG CCCTACACGC CAGGGCCCTA CCCTGGCCCT
501 GGGGGGCCA CAGCCACAGT CCTGGTCCCT TCAGGAGCTG CCACCACGGT
551 GACAGTGTCT CAGGGAGAGA TCTTTGAGGG AGCGCCTGTG CAGACGGTGT
601 GTCCCCACTG CCAGCAGGCC ATCGCCACCA AGATCTCCTA CGAGATTGGC
651 TTGATGAATT TCGTGCTGGG TTTCTTCTGT TGCTTCATGG GATGTGATCT
701 GGGCTGCTGC CTGATCCCTT GCCTCATCAA TGACTTCAAG GATGTGACGC
751 ACACATGCCC CAGCTGCAAA GCCTACATCT ACACGTACAA GCGCCTGTGC
801 TAACGGAGCT GGGACTCGGG ACTCCCCCGC CTGTCACTCT GGCCCCCTGT
851 GCTTTGCTCC CTGCGCTCAG TGGTCACTTT CCCGCTCCCA CTGGGGCTG
901 GGAGCCGTGC CACCATCCCC TAGAAGTCCT GTCCTCTTCA CCTGCCCCA
951 CCTGAGCCGC TGACTCTTCT GGCAAAAATT CTGTTGGGAT TTAAGGCCAA
1001 GGGTCAGTGG GTGGCAGGGG GCTGGCAATG AGCTTGTGTG TTGTTGGTCT
1051 GCTTGGTGTG TGTGATCGGG AAGATAAGCT GGGAGGGGTC TCCTGCTGGG
1101 GTCCTGATGC CTCTGTTTCC AAACAAGGTA CAGGTTCACT CCAGACTCTT
1151 TCCCCCTGGG ACCAACAGCA GCCAGAGCAG TTAGCCAGTT AGTCCCCAGG
1201 CCTGTGGCCA CAGGCGTTTC TGACCTGCTG GGCCGACAAT GGGTAAGTTG
1251 TCTGGAGTCA GGTGGGCCCA CGTAGGACAG GGTCACAAAG CCTGGGTTTG
1301 TTTCTGGGTA CTTGCGCCT CTGGGGTGCT AGAGGTGGGG CATGGTGGCT
1351 GGAAGTAAAA CTGCCAACTC TGGCCCTCAG AACTCTCAGG TATAGAAGCC
1401 CAGGATGTCT AATACCCTGT CCCAGTGCCC GAGAGCTGCC TGGTGTGAGG
1451 TAGAGAGGAC ACTGTACCTG GGTGAATGAT CAGACCTGAG TAGCTAAGAA
1501 GGAACCTGTC CCTTTGAGTC AGTGTGCAGA CCCCTTTTCA GGCCATGCCT
1551 CTGTGAACCC TGTATTGCTG GGGCCGGAAG GAGCCCTTGA GCCTAGCCCC
1601 TTCCCTCTCT CCCTGTGTCC TCACTGCGTG TGGGTATGAC CTCTGCCTGG
1651 TGGCTGTGTT ATCCCAACTG GGCAAGAGAT GGCAGAGGGT CCCCCTTGTG
1701 GGTGCGCTTG GATGTGCAGA GCCTTCTCCA TGGATTCTCT TCCCTGTAAG
1751 TGCCGGGGCC CCCACCCAG CTGACAGGCT GTTGCTGTGC CTGCTACAC
1801 CTGCTCCTCG AGGCACACTG GGCTAGGGAC GAGGAAGGAG CAGCCACAAG
1851 TGGTAGAACT GCCTTGGTGG ACACCAGCCT CGCCCTGTCT TTATTTCTG
1901 AATGGTTTGT GAACCTGTCT ACCTGGACCA CTGTATCCTG CCACTGTCTT
1951 TCCTGCTCTC GCACTGCCAC TGCATGGCCT CCTGTCACTG TGAATCGTGG
2001 CCCAGTCTCA GTTTGTAGTT TCTCATTAAT TTGGCCCTTT CACTCCCCCA
2051 AAAAAAAA
```

BLAST Results

Entry HS727347 from database EMBL:
 human STS WI-16589.
 Length = 275
 Plus Strand HSPs:
 Score = 1365 (204.8 bits), Expect = 3.0e-55, P = 3.0e-55
 Identities = 275/276 (99%), Positives = 275/276 (99%), Strand = Plus /
 Pl

Medline entries

No Medline entry

Peptide information for frame 3

```

1 MSSEPPPPYP GGPTAPLLEE KSGAPPTPGR SSPAVMQPPP GMPLPPADIG
51 PPPYEPPGHP MPQPGFIPPH MSADGTYMPP GFYPPPGPHP PMGYPPPGPY
101 TPGPYPGPGG HTATVLVPSG AATTVTVLQG EIFEGAPVQT VCPHCQQAIA
151 TKISYEIGLM NFVLGFFCCF MGCDLGCCLI PCLINDFKDV THTCPSCKAY
201 IYTYKRLC

```

ORF from 177 bp to 800 bp; peptide length: 208
 Category: similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_82g14, frame 3

PIR:S57447 HPBRII-7 protein - human, N = 1, Score = 206, P = 8.4e-16

PIR:A47655 spliceosome-associated protein SAP 62 - human, N = 1, Score
 = 198, P = 4.3e-15

>PIR:S57447 HPBRII-7 protein - human
 Length = 551

HSPs:

Score = 206 (30.9 bits), Expect = 8.4e-16, P = 8.4e-16
 Identities = 57/115 (49%), Positives = 62/115 (53%)

```

Query:      5 PPPPYPGGPTAPLLEEKSGAPPTPGRSSPAVMQPPPGMPLPPADIGPP-----PYEP--- 56
            PPPP+P G T P      G P PG      P      PPPG LPP GPP      P P
Sbjct:     226 PPPFFPAGQTPF--RPPLGPPGPPGPPGP----PPPGQVLPPLLAGPPNRRGDRPPPPVLF 279

Query:     57 PGHPMPQP--GFIPPHMSADGTYMP-PGFYPPPGPHPPM-GYYP-GPYTPGPYFGPGGH 111
            PG P QP G +PP      G      P PG+ PPPGP PP G PP GP+ P P PGP G
Sbjct:     280 PGQFPGQPPLGLLPP-----GPPPPVPGYGPPPGPPPPQOGPPPPPGPFPPRP-PGPLGP 333

Query:     112 TATVLVP 118
            T+  P
Sbjct:     334 PLTLAPP 340

```

Score = 177 (26.6 bits), Expect = 1.1e-12, P = 1.1e-12
 Identities = 55/120 (45%), Positives = 61/120 (50%)

```

Query:      5 PPPPYPGGPTAP--LLEEKSGAPPTPG-RSSPAVM---QP---PPGMPLPPADIGPPPYE 55
            P PP P GP P +L      PP G R P V+  QP PP PLPP GPPP
Sbjct:     244 PGPPGPPGPPPPGQVLPPLLAGPPNRRGDRPPPPVLFPGQFPGQPLGLLPP---GPPP-P 299

Query:     56 PGHPMPQPFGFIPPHMSADGTYMPPGFYPP--PGP-HPPMGYPPPGPYTPGPYPG---PG 109
            PG+ P PG PP      G      PPG +PP PGP PP+ PP P+ PGP PG P
Sbjct:     300 VPGYG-PPPGPPPPQ---GPPPPPGFPPRPFGPLGPPLTLAPP-PHLEGPPPGAPPFA 354

Query:     110 GHTATVLVP 118
            H      P
Sbjct:     355 PHVNPAFFP 363

```

Score = 168 (25.2 bits), Expect = 1.1e-11, P = 1.1e-11
 Identities = 47/118 (39%), Positives = 51/118 (43%)

Query: 5 PPPPYPG-GPTAPLLEEKSGAPPTPGRSSPAVMQP---PPGMPLPPADI-GPPPYEPPGHP 60

Score = 156 (23.4 bits), Expect = 2.1e-10, P = 2.1e-10
Identities = 44/103 (42%), Positives = 50/103 (48%)

Score = 121 (18.2 bits), Expect = 5.2e-05, P = 5.2e-05
Identities = 40/90 (44%), Positives = 45/90 (50%)

Pedant information for DKFZphfbr2_82g14, frame 3

Report for DKFZphfbr2_82g14.3

```
SEQ      PCLINDFKDVTHTCPSCAYIITYKRLC
SEG      .....
PRD      eeeeeccccccccccccceeeeeeeccc
MEM      MMMM.....
```

Prosites for DKFZphfbr2_82g14.3

(No Pfam data available for DKFZphfbr2_82g14.3)

DKFZphfbr2_82117

group: signal transduction

DKFZphtes2_82117 encodes a novel 334 amino acid protein with similarity to the plasma membrane substrate for the cAMP-dependent protein kinase.

The novel protein is a transmembrane protein with strong similarity to the phospholemmann protein, a membrane substrate for the cAMP-dependent protein kinase. It seems to serve as a chloride channel or as a chloride-channel regulator.

The new protein can find application in modulating/blocking cAMP-dependent protein kinase-dependent pathways.

similarity to plasma membrane substrate for cAMP-dependent protein kinase

complete cDNA, complete cds, EST hits
potential start at Bp 31 matches Kozak consensus PyNnatgG
might be a SODIUM/POTASSIUM-TRANSPORTING ATPASE
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="11: 920_E_12; 786_(A,H)_11; (797,802)_(E,H)_7"

Insert length: 1647 bp
Poly A stretch at pos. 1637, polyadenylation signal at pos. 1615

```
1 AGTCTCGGAG GGGACCGGCT GTGCAGACGC CATGGAGTTG GTGCTGGTCT
51 TCCTCTGCAG CCTGCTGGCC CCCATGGTCC TGGCCAGTGC AGCTGAAAAG
101 GAGAAGGAAA TGGACCCTTT TCATTATGAT TACCAGACCC TGAGGATTGG
151 GGGACTGGTG TTCGCTGTGG TTCTCTTCTC GGTGGGATC CTCTTATCC
201 TAAGTCGCAG GTGCAAGTGC AGTTTCAATC AGAAGCCCCG GGGCCCAGGA
251 GATGAGGAAG CCCAGGTGGA GAACCTCATC ACCGCCAATG CAACAGAGCC
301 CCAGAAAGCA GAGAACTGAA GTGCAGCCAT CAGGTGGAAG CCTCTGGAAC
351 CTGAGCGCGC TGCTTGAACC TTTGGATGCA AATGTCGATG CTTAAGAAAA
401 CCGGCCACTT CAGCAACAGC CCTTCCCCA GGAGAAGCCA AGAACTTGTG
451 TGTCCCCAC CCTATCCCT CTAACACCAT TCCTCCACCT GATGATGCAA
501 CTAACACTTG CCTCCCGCT GCAGCCTGTG GTCCTGCCCA CCTCCCGTGA
551 TGTGTGTGTG TGTGTGTGTG TGTGTGACTG TGTGTGTTT CTAAGTGTGG
601 TCTTTGTGGC TACTTGTGTT TGGATGGTAT TGTGTTTGT AGTGAAGTGT
651 GGACTCGCTT TCCCAGGCAG GGGCTGAGCC ACACGGCCAT CTGCTCTCTC
701 CTGCCCCCGT GGCCTCCAT CACCTTCTGC TCCTAGGAGG CTGCTTGTG
751 CCGGAGACCA GCCCCCTCCC CTGATTAGG GATGCGTAGG GTAAGAGCAC
801 GGGCAGTGGT CTTAGTCGT CTTGGGACCT GGAAGGTTT GCAGCACTTT
851 GTCATCATTC TTCATGGACT CCTTCACTC CTTAAGCAAA AACCTTGCTT
901 CCTATCCCA CTTGATCCCA GTCTGAAGGT CTCTTAGCAA CTGGAGATAC
951 AAAGCAAGGA GCTGGTGAGC CCAGCGTTGA CGTCAGGCAG GCTATGCCCT
1001 TCCGTGGTTA ATTTCTTCCC AGGGGCTTCC ACGAGGAGTC CCCATCTGCC
1051 CCGCCCCCTT ACAGAGCGCC CGGGGATTCC AGGCCAGGG CTTCTACTCT
1101 GCCCTGGGG AATGTGTCCC CTGCATATCT TCTCAGCAAT AACTCCATGG
1151 GCTCTGGGAC CCTACCCCTT CCAACCTTCC CTGCTTCTGA GACTTCAATC
1201 TACAGCCCAG CTCATCCAGA TGCAGACTAC AGTCCCTGCA ATTGGGTCTC
1251 TGGCAGGCAA TAGTTGAAGG ACTTCCTGTT CCGTTGGGGC CAGCACACCG
1301 GGATGGATGG AGGGAGAGCA GAGGCCTTTG CTTCTCTGCC TACGTCCCCT
1351 TAGATGGGCA GCAGAGGCAA CTCGCCATC CTTTGCTCTG CCTGTGAGTG
1401 GTCAGAGCGG TGAGCGAGGT GGGTTGGAGA CTCAGCAGGC TCCGTGCAGC
1451 CCTTGGGAAC AGTGAGAGGT TGAAGGTCAT AACGAGAGTG GGAAGTCAAC
1501 CCAGATCCCG CCCCTCCTGT CCTCTGTGTT CCCGCGGAAA CCAACCAAC
1551 CGTGCGCTGT GACCCATTGC TGTCTCTGT ATCGTGACCT ATCCTCAACA
1601 ACAACAGAAA AAAGGAATAA AATATCCTTT GTTTCCTAAA AAAAAAA
```

BLAST Results

Entry HS31455 from database EMBL:
human STS WI-2739.
Length = 103
Minus Strand HSPs:
Score = 487 (73.1 bits), Expect = 4.4e-14, P = 4.4e-14
Identities = 101/104 (97%), Positives = 101/104 (97%), Strand = Minus /
Plus
frame shift in primer binding site

Medline entries

91250422:
Purification and complete sequence determination of the major plasma membrane substrate for cAMP-dependent protein kinase and protein kinase C in myocardium.

95091702:
Protein kinase C and cyclic AMP-dependent protein kinase phosphorylate phospholemman, an insulin and adrenaline-regulated membrane phosphoprotein, at specific sites in the carboxy terminal domain.

95138184:
Mat-8, a novel phospholemman-like protein expressed in human breast tumors, induces a chloride conductance in *Xenopus* oocytes.

Peptide information for frame 2

1 MELVLVFLCS LLAPMVLASA AEKEKEMDPF HYDYQTLRIG GLVFAVVLFS
51 VGILLILSRK CKCSFNQKPR APGDDEEAQVE NLITANATEP QKAEN

ORF from 32 bp to 316 bp; peptide length: 95
Category: strong similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i17, frame 2

SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 196, P = 1.2e-15

TREMBL:AF091390.1 product: "phospholemman precursor"; Mus musculus phospholemman precursor, gene, complete cds., N = 1, Score = 187, P = 1.1e-14

PIR:A40533 cAMP-dependent protein kinase major membrane substrate precursor - dog, N = 1, Score = 189, P = 6.5e-15

SWISSPROT:PLM_RAT PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 185, P = 1.7e-14

>SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR.
Length = 92

HSPs:

Score = 196 (29.4 bits), Expect = 1.2e-15, P = 1.2e-15
Identities = 43/85 (50%), Positives = 56/85 (65%)

Query: 4 VLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSRRCCK 63
+LVF LL +AE KE DPF YDYQ+L+IGGLV A +LF +GIL++LSRRC+C
Sbjct: 7 ILVFCVGLLT----MAKAESPKEHDPFTYDYSLSLQIGGLVIAGILFILGILIVLSRRCRC 62

Query: 64 SFNQKPRA--PGDDEEAQVENLITANAT 88
FNQ+ R P +EE + I +T
Sbjct: 63 KFNQQQRTGEPDEEETFRSSIRRLST 89

Pedant information for DKFZphfbr2_82i17, frame 2

Report for DKFZphfbr2_82i17.2

[LENGTH] 95
[MW] 10542.37
[pI] 5.05
[HOMOL] SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR. 3e-15
[BLOCKS] BL01310

[EC]	3.6.1.37 Na ⁺ /K ⁺ -exchanging ATPase	6e-08
[PIRKW]	transmembrane protein	1e-09
[PIRKW]	hydrolase	6e-08
[PROSITE]	ATP1G1_PLM_MAT8	1
[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	1
[PROSITE]	TYR_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	2
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	
[KW]	SIGNAL PEPTIDE 19	

SEQ MELVLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFVAVLFSVGILLILSR
PRD cccchhhhhhhhhccccccccccccccccceeeeccccceehhhhhhheeeehhh

SEQ CKCSFNQKPAPGDEEAQVENLITANATEPQAEN
PRD hhccccccccccchhhhhhhhhhhcccccccc

Prosites for DKFZphfbr2_82i17.2

PS00001	86->90	ASN_GLYCOSYLATION	PDOC00001
PS00005	36->39	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00007	25->33	TYR_PHOSPHO_SITE	PDOC00007
PS00008	41->47	MYRISTYL	PDOC00008
PS01310	28->42	ATP1G1_PLM_MAT8	PDOC01014

(No Pfam data available for DKF2phfbr2_82i17.2)

DKFZphfbr2_82i24

group: nucleic acid management

DKFZphfbr2_82i24 encodes a novel 547 amino acid protein with similarity to DEAD-box superfamily ATP-dependent helicases.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis.

The novel protein contains a DEAD-box an ATP/GTP-binding site motif A (P-loop, interacting with one of the phosphate groups of the nucleotide) and a leucine zipper. Mutations in the closely related *Drosophila* Hlc gene result in lethality in homozygotes. Therefore the new protein seems to be critical involved in RNA processing in eukaryotic cells.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to DEAD-box subfamily ATP-dependent helicase

complete cDNA, complete cds, EST hits
potential Start at Bp 9 matches Kozak consensus PyNNatgG,
[PFAM] Helicases conserved C-terminal domain
[PFAM] DEAD and DEAH box helicases

Sequenced by DKFZ

Locus: /map="720_A_3; 758_H_4; 772_E_3; 804_A_5; 175.5 cR from topFT of Chr7 linkage group"

Insert length: 1860 bp

Poly A stretch at pos. 1850, polyadenylation signal at pos. 1829

```
1 AGCAGCGCCA TGGAGGACTC TGAAGCACTG GGCTTCGAAC ACATGGGCCT
51 CGATCCCCGG CTCCTTCAGG CTGTCACCGA TCTGGGCTGG TCGCGACCTA
101 CGGTGATCCA GGAGAAGGCC ATCCCCTGCG CCCTAGAAGG GAAGGACCTC
151 CTGGCTCGGG CCCGCACGGG CTCCGGGAAG ACGGCCGCTT ATGCTATTCC
201 GATGCTGCAG CTGTTGCTCC ATAGGAAGGC GACAGGTCCG GTGGTAGAAC
251 AGGCAGTGAG AGGCCTTGT CTGTTCCTA CCAAGGAGCT GGCACGGCAA
301 GCACAGTCCA TGATTACGCA GCTGGCTACC TACTGTGCTC GGGATGTCCG
351 AGTGGCCAAT GTCTCAGCTG CTGAAGACTC AGTCTCTCAG AGAGCTGTGC
401 TGATGGAGAA GCCAGATGTG GTAGTAGGGA CCCCATCTCG CATATTAAGC
451 CACTTGACAGC AAGACAGCCT GAAACTTCGT GACTCCCTGG AGCTTTTGGT
501 GGTGGACGAA GCTGACCTTC TTTTTCCTT TGGCTTTGAA GAAGAGCTCA
551 AGAGTCTCCT CTGTCACTTG CCCCGGATTT ACCAGGCTTT TCTCATGTCA
601 GCTACTTTTA ACGAGGACGT ACAAGCACTC AAGGAGCTGA TATTACATAA
651 CCCGTTTACC CTTAAGTTAC AGGAGTCCCA GCTGCCTGGG CCAGACCACT
701 TACAGCAGTT TCAGGTGGTC TGTGAGACTG AGGAAGACAA ATTCCTCCTG
751 CTGTATGCCC TGCTCAAGCT GTCATTGATT CGGGGCAAGT CTCTGCTCTT
801 TGTCAACACT CTAGAACGGA GTTACCGGCT ACGCTGTGTC TTGGAACAGT
851 TCAGCATCCC CACCTGTGTG CTCAATGGAG AGCTTCCACT GCGCTCCAGG
901 TGCCACATCA TCTCAGCTT CAACCAAGGC TTCTACGACT GTGTACATAGC
951 AACTGATGCT GAAGTCCTGG GGGCCCCAGT CAAGGGCAAG CGTCGGGGCC
1001 GAGGGCCCCA AGGGGACAAG GCCTCTGATC CGGAAGCAGG TGTGGCCCCG
1051 GGCATAGACT TCCACCATGT GTCTGCTGTG CTCAACTTTG ATCTTCCCCC
1101 AACCCTGAGG GCCTACATCC ATCGAGCTGG CAGGACAGCA CGCGCTAACA
1151 ACCCAGGCAT AGTCTTAACC TTTGTGCTTC CCACGGAGCA GTTCCACTTA
1201 GGCAAGATTG AGGAGCTTCT CAGTGGAGAG AACAGGGGCC CCATTCTGCT
1251 CCCCTACCAG TTCCGGATGG AGGAGATCGA GGGCTTCCGC TATCGCTGCA
1301 GGGATGCCAT GCGCTCAGTG ACTAAGCAGG CCATTCCGGA GGCAGATTG
1351 AAGGAGATCA AGGAAGAGCT TCTGCATTCT GAGAAGCTTA AGACATACTT
1401 TGAAGACAAC CCTAGGGACC TCCAGCTGCT GCGGCATGAC CTACCTTTGC
1451 ACCCCGAGT GGTGAAGCCC CACCTGGGCC ATGTTCTCTG CTACCTGGTT
1501 CCTCTGTGCT TCCGTGGCCT GGTACGCCCT CACAAGAAGC GGAAGAAGCT
1551 GTCTTCTCTT GTAGGAAGG CCAAGAGAGC AAAGTCCCAG AACCCACTGC
1601 GCAGCTTCAA GCACAAAGGA AAGAAATTCA GACCCACAGC CAAGCCCTCC
1651 TGAGGTTGTT GGGCCTCTCT GGAGCTGAGC ACATTGTGGA GCACAGGCTT
1701 ACACCCCTTC TGGACAGGCG AGGCTCTGGT GCTTACTGCA CAGCCTGAAC
1751 AGACAGTTCT GGGGCCGGCA GTGCTGGGCC CTTAGCTCC TTGGCACTTC
1801 CAAGCTGGCA TCTTGCCCTT TGACAACAGA ATAAAAATTT TAGCTGCCCC
1851 AAAAAA
```

BLAST Results

Entry HSG05793 from database EMBL:

human STS WI-6581.

Length = 206

Minus Strand HSPs:

Score = 992 (148.8 bits), Expect = 6.0e-38, P = 6.0e-38

Identities = 204/208 (98%), Positives = 204/208 (98%), Strand = Minus / P1

Entry AC004938 from database EMBL:

Homo sapiens clone DJ0971C03; HTGS phase 1, 18 unordered pieces.

Score = 1269, P = 6.5e-202, identities = 269/282

12 exons Bp ~87920-93706 (matching 1-1497)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 1650 bp; peptide length: 547

Category: strong similarity to known protein

Classification: Nucleic acid management

Prosite motifs: ATP_GTP_A (51-59)

LEUCINE_ZIPPER (149-171)

```

1 MEDSEALGFE HMGLOPRLLO AVTDLGWSRP TLIQEKAIPL ALEGKDLLAR
51 ARTGSGKTAA YAIPMLQLLL HRKATGPVVE QAVRGLVLVP TKELARQAQS
101 MIQQLATYCA RDVRVANVSA AEDSVSQRAV LMEKPDVVVG TPSRILSHLQ
151 QDSLKLROSL ELLVVDEADL LFSFGFEEEL KSLLCHLPRI YQAFILMSATF
201 NEDVQALKEL ILHNPVTLKL QESQLPGPDQ LQQFQVVCET EEDKFLLLYA
251 LLKLSLIRGK SLLFVNTLER SYRLRLFLEQ FSIPTCVLNG ELPLRSRCHI
301 ISQFNQGFYD CVIATDAEVL GAPVKGKRRG RGPKGDKASD PEAGVARGID
351 FHHVSAVLNF DLPPTPEAYI HRAGRTARAN NPGIVLTFVL PTEQFHLGKI
401 EELLSGENRG PILLPYQFRM EEIEGFYRC RDAMRSVTKQ AIREARLKEI
451 KEELLHSEKL KTYFEDNPRD LQLLRHDLPL HPAVVKPHLG HVPDYLVPFA
501 LRGLVRPHKK RKKLSSSCRK AKRAKSNPL RSFKHKGKKF RPTAKPS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i24, frame 1

TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds., N = 1, Score = 1230, P = 3.2e-125

TREMBL:SPCC1494_6 gene: "SPCC1494.06c"; product: "atp dependent helicase"; S.pombe chromosome II cosmid c1494., N = 2, Score = 753, P = 2.5e-113

PIR:S51412 hypothetical protein YLR276c - yeast (Saccharomyces cerevisiae), N = 2, Score = 711, P = 8.2e-117

TREMBL:AF025451_2 gene: "C24H12.4"; Caenorhabditis elegans cosmid C24H12., N = 2, Score = 564, P = 2.7e-99

>TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds.
Length = 560

HSPs:

Score = 1230 (184.5 bits), Expect = 3.2e-125, P = 3.2e-125
Identities = 251/497 (50%), Positives = 344/497 (69%)

Query: 9 FEHMGDPRLQAVTDLQWSRPTLIQEKAIPLALEGKDLLARARTGSGKTAAYAI PMLQL 68
F + LD R+L+AV LGW +PTLIQ AIPL LEGKD++ RARTGSGKTA YA+P++Q

Sbjct: 11 FHELELDQRILKAVAQLGWQOFTLIQSTAIPLLEGGKDVVVRARTGSGKTATYALPLIQK 70

Query: 69 LLHRKATGPVVEQAVRGLVLPVKELARQAQSMIQQALATYCARDVRVANVS-AAEDSVSQ 127
+L+ K EQ V +VL PTKEL RQ++ +I+QL C + VRVA+++ ++ D+V+Q

Sbjct: 71 ILNSKLNAS--EQYVSAVVLAPTKELCRQSRKVIEQLVESCGKVVRVADIADSSNDTVTQ 128

Query: 128 RAVLMEKPDVVVGTPSRILSHLQDQSLKLRDSLELLVVDEADLLFSFGFEEELKSLCHL 187
R L E PD+VV TP+ +L++ + S+ +E LVVDEADL+F++G+E++ K L+ HL

Sbjct: 129 RHALSESPDIVVATPANLLAYAEAGSVVDLKHVETLVVDEADLVFAYGYEKDFKRLIKHL 188

Query: 188 PRIYQAFMSATFNEDVQALKEILHNVPVTLKLQESQLPGPDQLQFQVVCETEEDKFL 247
P IYQA L+SAT +DV +K L L+NPVTLKL+E +L DQL +++ E E DK +

Sbjct: 189 PPIYQAVLVSATLTDDVVRMKGCLNNPVTLEKEPELVQDQLSHQRILAE-ENDKPAI 247

Query: 248 LYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQ 307
LYALLKL LIRGKS++FVN+++R Y++RLFLEQF I CVLN ELP R H ISQFN+G

Sbjct: 248 LYALLKLRLIRGKSIIFVNSIDRCYKVRLEQFGIRACVLNSEL PANIRIHTISQFNKG 307

Query: 308 FYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPE 367
YD +IA+D + P G + K ++ D E+ +RGIDF V+ V+NFD P

Sbjct: 308 TYDIIIASDEHMEKP--GGKSATNRKSPRSMDMESSASRGIDFQCVNNVINFDFPRDVT 365

Query: 368 AYIHRAGRTARANNPGIVLTFVLPTQFHLGKIEELL---SGENRGPIILPYQFRMEEI 423
+YIHRAGRTAR NN G VL+V E +E+ L + + I+ YQF+MEE+

Sbjct: 366 SYIHRAGRTARGNNKGSVLSFVSMKESKVNSVEKKLDCSFAAQEGEQIKNYQFKMEEV 425

Query: 424 EGFYRRCRDAMRSVTKQAIAREARLKEIKEELLHSEKLKTYFEDNPRDLQLLRHDLPLHPA 483
E FRYR +D R+ T+ A+ + R++EIK E+L+ EKLK +FE+N RDLQ LRHD PL

Sbjct: 426 ESFRYRAQDCWRAATRVAVHDTRIIEIKIEILNCEKLEKAFEEENKRDQLALRHDKPLRAI 485

Query: 484 VVKPHLGHVDPDYLVPALRGLV 505
V+ HL +P+Y+VP AL+ +V

Sbjct: 486 KVQSHLSDMPEYIVPKALKRVV 507

Pedant information for DKF2phfbr2_82i24, frame 1

Report for DKF2phfbr2_82i24.1

[LENGTH] 547
[MW] 61589.88
[pI] 9.34
[HOMOL] TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster
tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent
bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst),
and la costa (lcs) genes, complete cds. 1e-121
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 1e-109
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA]
2e-42
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YLL008w] 8e-40
06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S.
cerevisiae, YKR059w] 3e-39
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YKR059w] 3e-39
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 3e-35
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 3e-29
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-29
[FUNCAT] 1 genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 1e-27
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-27
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-21
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 1e-05
[BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW] nucleus 4e-34
[PIRKW] RNA binding 7e-41
[PIRKW] DEAD box 2e-38
[PIRKW] transmembrane protein 9e-20
[PIRKW] DNA binding 8e-23
[PIRKW] ATP 1e-107
[PIRKW] purine nucleotide binding 2e-38
[PIRKW] P-loop 1e-107
[PIRKW] hydrolase 2e-35
[PIRKW] protein biosynthesis 2e-38
[PIRKW] ATP binding 7e-43


```

[SUPFAM]      WW repeat homology 1e-26
[SUPFAM]      DEAD/H box helicase homology 1e-107
[SUPFAM]      unassigned DEAD/H box helicases 1e-107
[SUPFAM]      ATP-dependent RNA helicase DBP1 3e-31
[SUPFAM]      ATP-dependent RNA helicase DHH1 2e-35
[SUPFAM]      translation initiation factor eIF-4A 2e-38
[SUPFAM]      tobacco ATP-dependent RNA helicase DB10 1e-26
[PROSITE]     ATP_GTP_A 1
[PROSITE]     LEUCINE_ZIPPER 1
[PFAM]        Helicases conserved C-terminal domain
[PFAM]        DEAD and DEAH box helicases
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY 9.87 %

```

```

SEQ      MEDSEALGFEHMGDLPRLLQAVTDLGWSRPTLIQEKAIPALLEGKDLLARARTGSGKTAA
SEG      .....
PRD      cccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ      YAIPMLQLLHRKATGPVVEQAVRGLVLPVKELARQAQSMIQLATYCARDVRVANVSA
SEG      .....
PRD      ehhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhccceeeeeecc

SEQ      AEDSVSQRAVLMEKPDVVVGTPSRILSHLQQDSLKLRLDSLELLVVDEADLLFSFGFEEEL
SEG      .....
PRD      ccchhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhccchhhh

SEQ      KSLLCCHLPRIYQAFMSATFNEDVQALKELILHNPVTLKLOESQLPGPDQLQFQVVCET
SEG      .....
PRD      hhhhhhhccchhhhhhhhhccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhh

SEQ      EEDKFLLLYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHI
SEG      .....
PRD      hhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      ISQFNQGFYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNF
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      DLPPTPEAYIHRAGRTARANNPGIVLTFVLPTQFHLGKIEELLSGENRGPILLPYQFRM
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ      EEIEGFRYRCRDAMRSVTKQAIAREARLKEIKEELLHSEKLKTYFEDNPRDLQLLRHDLPL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhcc

SEQ      HPAVVKPHLGHVPDYLVPALRGLVRPHKKRKKLSSSSCRKAKRAKSNPLRSFKHKGKKF
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhcccccccccccccccccc

SEQ      RPTAKPS
SEG      .....
PRD      ccccccc

```

Prosite for DKFZphfbr2_82124.1

PS00017	51->59	ATP_GTP_A	PDOC00017
PS00029	149->171	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_82124.1

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACaQTGSGKTAaF		
Query	13	GLDPRLQAVTDLGWSRPTLIQEKAIPALLEGKDLLARARTGSGKTAAY	61
HMM	1IPMLQHIdwdP...WpqpPQdPrALILAPTRELAMQIEEcRkFgkHmN		
Query	62	AIPMLQLLHRKATGPVVEQA-VRGLVLPVKELARQAQSMIQLATYCA	110
HMM	g.IRImCIYGGtnMRdQMRmLeRGpPHIViATPGRLIDHIERgtldLDr.		
Query	111	RDVRVANVSAEDSVSQRAVLMEKP-DVVVGTPSRILSHLQQDSLKLRLDS	159
HMM	IeMLVMDEADRMLDMGFIDQIRrImrqIPmpwnRQTMMSATMPdeIqEL		
	+E LV DEAD +++ GF++++ ++ ++P + Q + SAT+ +++Q L		


```

Query      160 LELLVVDEADLLFSFGFEEELKSLCHLP--RIYQAFILMSATFNEDVQAL  207
HMM        ARrFMRNPiRInIdMdElTtnEnikQwyiyVerEMWkfdclcrLle*
          + +++NP+ + + +++L + ++Q+ +++E E++KF +L+ L++
Query      208 KELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKFLLLYALLK  253

HMM_NAME   Helicases conserved C-terminal domain

HMM        *EilleeWLknIGirvmYIHGdMpQeERdeIMddFNnGEynVLicTDV...
          +L+ +L++ I+++++ G +P + R I+ +FN+G Y++ I+TD+
Query      272 YRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQGFYDCVIATDAEVL  320
HMM        .....ggRGIDIPdVNVHVINYDMPWNPEqYI
          +RGID+ V+ V N+D+P +PE YI
Query      321 GAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPEAYI  370
HMM        QRIGRTgRIG*
          +R+GRT+R++
Query      371 HRAGRTARAN  380

```


DKFZphfbr2_82m16

group: brain derived

DKFZphfbr2_82m16 encodes a novel 289 amino acid protein with very weak similarity to A.thaliana F28A23.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F28A23.140

complete cDNA, complete cds, few EST hits
many ATGs in front of the ORF
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="4"

Insert length: 2715 bp

Poly A stretch at pos. 2705, polyadenylation signal at pos. 2687

```
1 AGAGGAGGGG AGAGGACTGG GGAGCCGAGC CAGAGCCGGG CTGCCTGCCA
51 CCGGGCTGCT CGTCCGCTAG CTGGGGAGGA GCGCTCCACC CGCAACTGAC
101 AAAGGATGGG AGAATGCCCG CGCCCCGGGA TGCCGGCCGC ACGCAGCCTG
151 CCGGCCCGCT GAGCTACTTC ACCCTCCGCC GGTAAGTGAC TGCAAACATC
201 ATTCAATCAA TCAGCCTCAC TGGGAGCCCC TTCTCTCCGG CTGGTAGTCC
251 TGGGCGGCTT GTCCCTGATC CCGAGCGGGG CTTGGCACAG CATCAGCCCT
301 GGAGGGCAGG CAGCAGGTGC CTTTGCCTGG TGGGTCCACT GGGGAGCGTG
351 GCTGGGGTTC GCGGCGGGTG CTGCCACCCA ACCTGCGGGC GGCGGGCTCG
401 CCCAGTAGGC GCCTCTCTGG TGAGAGGAGG CGGCTCCAGC CCGCATCCTG
451 GGGTAGTTGC TACTATTGGC CCCCAGCGCC CGCTCTGCGC GCGCGCCGTT
501 TCTGGCGGAT CCCCAGTGCG CGGCGCGCTG TTTACACCGG CGTGGTACTA
551 GTCACGGAGC CGCACCCCTC GGAAAGCGCG GAGTCGATGA CAGCCACTTC
601 ACAGGCTCAC GCGCTCCTAG TGTGGGCTTG AAGGGGACGG GGACCGATTA
651 CCAAAGGAGA GCGCTGAGTA CGGAAGACAC AGGGCAGCCT TTGTCTTGGG
701 TTTAGCGCTG ATGCGCTCAA CCCTGAGTCG GGTTCAGTGC AACTGTTGTG
751 TCCGATTTTC GTTCCCTGCA ACCGCCCTCC TGGGCGAGAG ATGTCATTGT
801 GTTCTGTGGG CCAGCGGGAC TGAGAGCTGG GACTTAAGAC GCCAGGAGGG
851 TCCTGCGCTC ACGGGAAATG TACCCCAAAA GAATCTGAG AGAATATACT
901 CAACTGTCTT GCTGTGATTA AACAAGACTG CTGTATTTTA ATTTCAAAA
951 TTGAAAAGGG ATAGGAGGAA GGGGAAATG CTGGGCTGGT GTGAAGCGAT
1001 AGCCCGTAAC CCTCACAGAA TTCCAAACAA CACGCGAACA CCGGAGATCT
1051 CAGGGGATTT GGCTGACGCC TCACAAACCT CCACATTGAA TGAATAATCC
1101 CCAGGGCGAT CTGCAAGTCG ATCAAGTAAC ATTTCAAAG CAAGCAGCCC
1151 AACACAGGGG ACAGCTCCCA GGAGCCAGTC AAGGTTGTCT GTCTGTCCAT
1201 CCACTCAGGA CATCTGCAGA ATCTGTCACT GCGAAGGGGA TGAAGAGAGC
1251 CCCCTCATCA CACCCTGTGC CTGCACTGGG ACACCTGCGT TGTGCCACCA
1301 GTCCTGCCTC CACCACTGGA TAAAGAGTTC AGATACACGC TGCTGTGAGC
1351 TCTGCAAGTA TGACTTCATA ATGGAGACCA AGCTCAAACC CCTCCGGAAG
1401 TGGGAGAAAC TACAGATGAC CACAAGTGAA AGGAGGAAAA TATTCTGCTC
1451 TGTCACATTC CACGTAATCG CGATCACCTG TGTGGTTTGG TCTTTGTATG
1501 TATTGATAGA CCGGACAGCG GAGGAAATCA AGCAAGGCAA TGACAATGGT
1551 GTCCTTGAAT GGCCATTTTG GACAAAATCG GTTGTGGTAG CCATTGGCTT
1601 CACAGGAGGT CTTGTCTTCA TGTACGTACA GTGTAAGTC TATGTTCACT
1651 TGTGGCGCAG GCTGAAGGCC TACAACCGTG TGATCTTTGT ACAAAATTGC
1701 CCAGACACTG CCAAAAACCT GGAGAAGAAC TTCTCATGTA ATGTAAACAC
1751 AGACATCAAA GATGCTGTGG TAGTGCCCTG ACCACAAACA GGTGCAAAAT
1801 CACTGCCATC TGCAGAGGGT GGCCCCCTG AAGTTGTATC AGTCTGATGG
1851 AACCTGTTGG GAGTTTCTTC ACCGAAGAAT ATCTTTCTAG CCTCAGCCA
1901 CTACAAATGA CAGAAGTGAC CTTGAATTAT TTAATCCTT CAGTCTCTCC
1951 TTTCTCCTAC TGACACATTT TTCTGACTT TGTTCAAAGA GGAAGGAGA
2001 AAAACAAACA AACAGACCAA ATGCCAGGGA GCCCATGAAG TAATAGCGTA
2051 AAGTAAAGTA TGATATGGAA ATGTGAAGTT TGCAAGAGAA TGATTTCCAA
2101 GACAATTAAG AACTACTGGG GCAATGAATG CTTTTAGGCA GTAATCAAG
2151 ATTAATGGGA CCCATGATAC TCTTCTTAC AGTAACAGGG GAAAGATTCA
2201 AGAATACAGA CTTGAATTGC GATGTGTATT ACTTCTAGGG CCTGTGAATG
2251 TTAATGTCTC CATCTGGAAA TAATACTAA CATATTGGT TTTAAGCCTG
2301 AAATTGTCTG CATTATCCCT AAGTCACATT GGAAGTGAAC TTGGAGGATG
2351 CATATTTTGA TATGCTTTGA CAGCTAACAG ATTTGTATGG TTTAGTGGAG
2401 TCTGGTTATT TTGACAGATG CATGTTTTTT TTAATAGAT GCAATATACA
2451 TTTGAAGACA TTGATATTTG GAATTAATTA TGTTTGTTTA AGTCACGCAA
2501 AAGATTTTCA GAAAATGTTT GGATATAATT AGCTCTGTTA AATACCCACA
2551 GAACTGTTAT CAGGTCTTAT ATTTATTTTC ATCTGGTTCC TCTAATACAG
```


2601 TGCTGTCCAA TAGAAACACA ACAGCCACAA ATGCAGGCCA CAGATGCAAA
 2651 TATTAACTT CCCAGTAGCC CTATTTTAAA AAGTAAAAAT AAATGTTTGT
 2701 TTGTTAAAAA AAAAA

BLAST Results

Entry G37457 from database EMBLNEW:
 SHGC-57357 Human Homo sapiens STS genomic.
 Length = 458
 Plus Strand HSPs:
 Score = 2116 (317.5 bits), Expect = 4.3e-91, P = 4.3e-91
 Identities = 444/456 (97%)

Medline entries

No Medline entry

Peptide information for frame 3

1 MLGWCEAIAR NPHRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS
 51 NISKASSPTT GTAPRSQSRL SVCSTQDIC RICHCEGDEE SPLITPCRCT
 101 GTLRFVHQSC LHQWIKSSDT RCCELCKYDF IMETKLKPLR KWEKLQMTTS
 151 ERRKIFCSVT FHVIAITCVV WSLYVLIDRT AEEIKQGNND GVLEWPFWTK
 201 LVVVAIGFTG GLVFMVQCK VYVQLWRLK AYNRVIFVQN CPDTAKKLEK
 251 NFSCNVNTDI KDAVVVPVPQ TGANSLPSAE GGPPEVVS

ORF from 978 bp to 1844 bp; peptide length: 289
 Category: similarity to unknown protein

BLASTP hits

Entry AB011169_1 from database TREMBL:
 gene: "KIAA0597"; product: "KIAA0597 protein"; Homo sapiens mRNA for
 KIAA0597 protein, partial cds.
 Score = 188, P = 6.0e-12, identities = 30/54, positives = 38/54

Entry SPBC14F5_7 from database TREMBL:
 gene: "SPBC14F5.07"; product: "hypothetical protein"; S.pombe
 chromosome II cosmid c14F5.
 Score = 185, P = 1.9e-11, identities = 29/53, positives = 38/53

Entry CEY57A10B_1 from database TREMBL:
 gene: "Y57A10B.1"; Caenorhabditis elegans cosmid Y57A10B
 Score = 171, P = 2.6e-10, identities = 40/107, positives = 58/107

Alert BLASTP hits for DKFZphfbr2_82m16, frame 3

TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII
 project), N = 1, Score = 198, P = 3.4e-13

>TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII project)
 Length = 1,051

HSPs:

Score = 198 (29.7 bits), Expect = 3.4e-13, P = 3.4e-13
 Identities = 38/103 (36%), Positives = 61/103 (59%)

Query: 28 LADASQTSTLNEKSPGRSASRS-SNISKASSPTTGTAPRSQSRLSVCSTQDICRICHCE 86
 +++ S +S+ + SP +++ SN+ A S TG+ +D+CRIC
 Sbjct: 20 VSEPSVSSSSSSSPNQASPNPFSNMDPAVSTATGSRYVDDE-----DEEDVCRICRNP 74

Query: 87 GDEESPLITPCRCTGTLRFVHQSC LHQWIKSSDTRCCELCKYDF 130
 GD ++PL PC C+G+++FVHQ CL QW+ S+ R CE+CK+ F
 Sbjct: 75 GDADNPLRYPACSGSIKFVHQDCLLQWLNHSNARQCEVCKHPF 118

Additional information for the proposed rulemaking:

Report for DKFZphfbr2 82m16.3

```
[LENGTH]          289
[MW]              32308.36
[pI]              8.76
[HOMOL]           PIR:T00268 hypothetical protein KIAA0597 - human (fragment) 9e-14
[FUNCAT]          04.99 other transcription activities [S. cerevisiae, YIL030c] 4e-09
[PIRKW]           transmembrane protein 9e-08
[PROSITE]         MYRISTYL 1
[PROSITE]         CK2_PHOSPHO_SITE 4
[PROSITE]         TYR_PHOSPHO_SITE 1
[PROSITE]         PKC_PHOSPHO_SITE 3
[PROSITE]         ASN_GLYCOSYLATION 3
[KW]              Alpha_Beta
[KW]              LOW_COMPLEXITY 6.57 %
```

```
SEQ      MLGWCEAIARNPHRI PNNTRTP EISGLDASQTSTLNEKSPGRSASRSSNISKASSPTT
SEG      .....XXXXXXXXXXXXXXXXXXXXX..
PRD      cccchhhhhccccccccccccccccccccchhhhhhhcccccccccccccccccccccccc
```

```
SEQ      GTAPRSQSRLSVC PSTQDICRICHCEGDEESPLITPCRCTGTLRFVHQSLHQQWIKSSDT  
SEG      .....  
PRD      cccccccccccccccceeeeecccccccccccccccccccceeehhhhhnhhhcc
```

```
SEQ      RCCELCKYDFIMETKLKPLRKWEKLQMTTSERRKIFCSVTFHVIAITCVVWSLYVLIDRT
SEG      .....
PRD      cccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
```

```

SEQ      AEEIKQGNONGVLEWPFWTKLVVVAIGFTGGLVFMVQCKVYVQLWRRLLKAYNRVIFVQN
SEG      .....
PRD      cccccccccccceehhhhhheeeeecccccceeeehhhhhhhhhhhhhhhheeeee

```

```
SEQ      CPDTAKKLEKNFSCNVNTDIKDAVVVPVPTGANSLPSEAEGGPPEVVS
SEG      .....
PRD      cccchhhhhccccccccccccceeeeecccccccccccccccccccc
```

Prosites for DKFZphfbr2_82m16.3

PS000001	17->21	ASN_GLYCOSYLATION	PDOC000001
PS000001	51->55	ASN_GLYCOSYLATION	PDOC000001
PS000001	251->255	ASN_GLYCOSYLATION	PDOC000001
PS000005	102->105	PKC_PHOSPHO_SITE	PDOC000005
PS000005	150->153	PKC_PHOSPHO_SITE	PDOC000005
PS000005	244->247	PKC_PHOSPHO_SITE	PDOC000005
PS000006	36->40	CK2_PHOSPHO_SITE	PDOC000006
PS000006	75->79	CK2_PHOSPHO_SITE	PDOC000006
PS000006	148->152	CK2_PHOSPHO_SITE	PDOC000006
PS000006	180->184	CK2_PHOSPHO_SITE	PDOC000006
PS000007	121->129	TYR_PHOSPHO_SITE	PDOC000007
PS000008	187->193	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2_82m16.3)

DKFZphfbr2_82m6

group: signal transduction

DKFZphfbr2_82m6.3 encodes a novel 654 amino acid protein with similarity to murine sphingosine kinase.

Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellularly, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependend on SPP. Extracellularly, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1.

The new protein can find application in modulating/blocking the shingosine kinase intracellular signal transmission pathway.

strong similarity to mouse "sphingosine kinase"

complete cDNA, complete cds, EST hits,
YLR260w/YOR171c Lcb5p/Lcb4p = long chain base kinases,
involved in biosynthesis of sphingolipids

Sequenced by DKFZ

Locus: unknown

Insert length: 2875 bp

Poly A stretch at pos. 2865, polyadenylation signal at pos. 2838

```

1 AGTGTGGAG GTGAGGAGGC GGGGCTGGCA GGGCTAGTCG GGGCATCTGG
51 AAATTTCCGA CCCACGCTT CGGGCGTTTC CTTATCAGGT TCACCGCTCC
101 CTGATCTCGC GCTGCACTTC GTAGGCGCAG CCGCTGCTTG GGAAGTCCTA
151 CTTAAGAGCT GAAGGTACAG CCAGGACAGT GAGACCTGAC TCCTTGCTCC
201 TACCAGCCTA CTATGGCTTA AGACCCAGGG CCAGGGTCCC GTTGATGTAA
251 CAGAGCAGAG GACCAGCAGA TGAATGGACA CCTTGAAGCA GAGGAGCAGC
301 AGGACCAGAG GCCAGACCAG GAGCTGACCG GGAGCTGGGG CCACGGGCCT
351 AGGAGCAGCC TGGTCAGGGC TAAGGCCATG GCCCGGCCCC CACCGCCACT
401 GGCTGCCAGC ACCTCGCTCC TCCATGGCGA GTTTGGCTCC TACCCAGCCC
451 GAGGCCACAG CTTTGCCCTC ACCCTTACAT CGCAGGCCCT GCACATACAG
501 CGGCTGCCGC CCAAACCTGA AGCCAGGCCC CGGGGTGGCC TGGTCCCGTT
551 GGCCGAGGTC TCAGGCTGCT GCACCTGCGC AAGCCGACG CCCTCAGACT
601 CAGCGGCCTA CTTCTGCATC TACACCTACC CTCGGGGCCG GCGCGGGGCC
651 CGGCGCAGAG CCACTCGCAC CTTCCGGGCA GATGGGGCCG CCACCTACGA
701 AGAGAACCCT GCCGAGGCCC AGCGCTGGGC CACTGCCCTC ACCTGTCTGC
751 FCCGAGGACT GCCACTGCCC GGGGATGGGG AGATCACCCC TGACCTGCTA
801 CCTCGGCCCG CCCGTTGCT TCTATTGGTC AATCCCTTTG GGGGTGGGG
851 CCTGGCCTGG CAGTGGTGTA AGAACCACGT GCTTCCCATG ATCTCTGAAG
901 CTGGGCTGTC CTTCAACCTC ATCCAGACAG AACGACAGAA CCACGCCCGG
951 GAGCTGGTCC AGGGGCTGAG CCTGAGTGAG TGGGATGGCA TCGTCACGGT
1001 CTCGGGAGAG GGGCTGCTCC ATGAGGTGCT GAACGGGCTC CTAGATCGCC
1051 CTGACTGGGA GGAAGCTGTG AAGATGCCTG TGGGCATCCT CCCCTGCGGC
1101 TCGGGCAACG CGCTGGCCGG AGCAGTGAAC CAGCACGGGG GATTTGAGCC
1151 AGCCCTGGGC CTCGACCTGT TGCTCAACTG CTCACTGTTG CTGTGCCGGG
1201 GTGGTGGCCA CCCACTGGAC CTGCTCTCCG TGACGCTGGC CTCGGGCTCC
1251 CGCTGTTTCT CTTTCTGTC TGTGGCCTGG GGCTTCGTGT CAGATGTGGA
1301 TATCCAGAGC GAGCGCTTCA GGGCCTTGGG CAGTGCCCGC TTCACACTGG
1351 GCACGGTGCT GGGCCTCGCC ACACCTGCAC CCTACCGCGG ACGCCTCTCC
1401 TACCTCCCCG CCACTGTGGA ACCTGCCTCG CCCACCCCTG CCCATAGCCT
1451 GCCTCGTGCC AAGTCGGAGC TGACCCTAAC CCCAGACCCA GCCCGGCCCA
1501 TGGCCCACTC ACCCCTGCAT CGTTCGTGT CTGACCTGCC TCTTCCCTG
1551 CCCCAGCCTG CCCTGGCCTC TCCTGGCTCG CCAGAACCCC TGCCCATCCT
1601 GTCCCTCAAC GGTGGGGGCC CAGAGCTGGC TGGGACTGG GGTGGGGCTG
1651 GGGATGCTCC GCTGTCCCG GACCCACTGC TGTCTTACC TCCTGGCTCT
1701 CCCAAGGCAG CTCTACACTC ACCCGTCTCC GAAGGGGCCC CCGTAATTCC
1751 CCCATCCTCT GGGCTCCAC TTCCACCCC TGATGCCCCG GTAGGGGCCT
1801 CCACCTGCGG CCGGCCGAC CACCTGCTGC CTCGCTAGG CACCCGCTG
1851 CCCCAGACT GGGTGACGCT GGAGGGGGAC TTTGTGCTCA TGTGGCCAT
1901 CTCGCCCAGC CACCTAGGCG CTGACCTGGT GGCAGCTCCG CATGCGCGCT
1951 TCGACGACGG CCTGGTGCAC CTGTGCTGGG TGGCTAGCGG CATCTCGCGG
2001 GCTGCGCTGC TGGCCTTTT CTTGGCCATG GAGCGTGTA GCCACTCAG
2051 CCTGGGCTGT CCGCAGCTGG GCTACGCCGC GGCCCGTGCC TTCCGCCCTAG
2101 AGCCGCTCAC ACCACGCGGC GTGCTCACAG TGGACGGGGA CGAGGTGGAG
2151 TATGGGCCGC TACAGGCACA GATGCACCCT GGCATCGGTA CACTGCTCAC
2201 TGGGCCCTCT GGCTGCCCCG GCGGGGAGCC CTGAACTAA ACAAGCTTGG
2251 TACCCGCGCG GGGCGGGGCC TACATTCCAA TGGGGCGGAG CCTGAGCTAG
2301 GGGGTGTGGC CTGGCTGCTA GAGTTGTGGT GGCAGGGGCC CTGGCCCGT

```



```

2351 CTCAGGATTG CGCTCGCTTT CATGGGACCA GACGTGATGC TGAAGGTGG
2401 GCGTCGTCAC GGTAAAGAG AAATGGGCTC GTCCCGAGGG TAGTGCCTGA
2451 TCAATGAGGG CGGGCCCTGG CGTCTGATCT GGGGCCGCC TTACGGGGCA
2501 GGGCTCAGTC CTGACGCTTG CCACCTGCTC CTACCCGGCC AGGATGGCTG
2551 AGGGCGGAGT CTATTTTACG CGTCGCCCAA TGACAGGACC TGAATGTAC
2601 TGGCTGGGCT AGGCCTCAGT GAGTCGGCCG GTCAGGGCCC GCAGCCTCGC
2651 CCCATCCACT CCGGTGCCTC CATTTAGCTG GCCAATCAGC CCAGGAGGGG
2701 CAGGTTCCCC GGGGCCGCGC CTAGGATTG CACTAATGTT CCTCTCCCCG
2751 CGGGTGGGGG CGGGGAAATT CATATCCCCT GTTCGTCTCA TGCAGTCCT
2801 CCGTCCCAA TCTAAAAAGC AATTGAAAAG GTCTATGCAA TAAAGGCAGT
2851 CGTTTCATTC CTCTCAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

99045661:

Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway.

98395082:

Molecular cloning and functional characterization of murine sphingosine kinase.

98241633:

Purification and characterization of rat kidney sphingosine kinase.

99178622:

Sphingosine 1-phosphate: a prototype of a new class of second messengers.

Peptide information for frame 3

```

1 MNGHLEAEEQ QDQRPDQELT GSWGHGPRST LVRAKAMAPP PPPLAASTSL
51 LHGEFGSYPA RGPRFALTLT SQALHIQRLR PKPEARPRGG LVPLAEVSGC
101 CTLRSRSPSD SAAYFCIYTY PRGRRGARRR ATRTFRADGA ATYEENRAEA
151 QRWATALTCL LRGLPLPGDG EITPDLLPRP PRLLLLVNPF GGRGLAWQWC
201 KNHVLPmise AGLSFNLIQT ERQNHARELV OGLSLSEWDG IVTVSGDGLL
251 HEVLNGLDDR PDWEEAVKMP VGILPCGSGN ALAGAVNQHG GFEPALGLDL
301 LNCSSLCLCR GGGHPLDLLS VTLASGSRCF SFLSVWGFV SDVDIQSERF
351 RALGSARFTL GTVLGLATLH TYRGRLSYLP ATVEPASPTP AHSLEPRKSE
401 LTLTPDPAPP MAHSPLHRSV SDLPPLPQP ALASPGSPPEP LPILSLNGGG
451 PELAGDWGGA GDAPLSPDPL LSSPPGSPKA ALHSPVSEGA PVIPPSSGLP
501 LPTPDARVGA STCGPPDHL L PPLGTPLPPD WVTLEGDFVL MLAISSPHLG
551 ADLVAAPHAR FDDGLVHLCW VRSGISRAAL LRLFLAMERG SHFSLGCPQL
601 GYAAARAFRL EPLTPRGVLT VDGEQVEYGP LQAQMHPGIG TLLTGPPGCP
651 GREP

```

ORF from 270 bp to 2231 bp; peptide length: 654
Category: similarity to known protein

BLASTP hits

Entry SPAC4A8_7 from database TREMBL:

gene: "SPAC4A8.07c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c4A8.

Score = 301, P = 7.9e-32, identities = 68/190, positives = 109/190

Entry CEC34C6_3 from database TREMBLNEW:

product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

>TREMBL:CEC34C6_3 product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

Score = 273, P = 9.0e-29, identities = 78/265, positives = 142/265

Entry S67059 from database PIR:

hypothetical protein YOR171c - yeast (Saccharomyces cerevisiae)

>TREMBL:SC55021_9 gene: "O3615"; product: "O3615p"; Saccharomyces cerevisiae cosmid pUOA1258 from chromosome 15R. >TREMBL:SCYOR170W_2

S.cerevisiae chromosome XV reading frame ORF YOR170W

Score = 253, P = 2.0e-25, identities = 70/234, positives = 116/234

Entry S51398 from database PIR:

hypothetical protein YLR260w - yeast (*Saccharomyces cerevisiae*)
 >TREMBL:SCL8479_4 gene: "YLR260W"; product: "Ylr260wp"; *Saccharomyces cerevisiae* chromosome XII cosmid 8479.
 Score = 251, P = 1.0e-24, identities = 62/198, positives = 103/198

Alert BLASTP hits for DKFZphfbr2_82m6, frame 3

TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1b) mRNA, complete cds., N = 2, Score = 615, P = 1.2e-92

TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds., N = 2, Score = 616, P = 2e-92

TREMBL:ATF18E5_16 gene: "F18E5.160"; product: "putative protein"; *Arabidopsis thaliana* DNA chromosome 4, BAC clone F18E5 (ESSAII project), N = 2, Score = 370, P = 6.8e-33

>TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds.
 Length = 504

HSPs:

Score = 616 (92.4 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
 Identities = 128/260 (49%), Positives = 173/260 (66%)

Query: 154 ATALTCLLRGLPLPGDGEITPDLLPRPPRLLLLVNPFGGRLAWQWCKNHVLPIMISEAGL 213
 A C L + E LLPRP R+L+L+NP GG+G A Q ++ V P + EA +
 Sbjct: 110 APVAPCQREPRDLAMEFECPRGLLPRPCRVLLNPFQGGKQKALQLFQSRVQPFLEEAEI 169

Query: 214 SFNLIQTERQNHARELVQGLSLSEWDGIVTVSGDGLLHEVLNGLLDRPDWEEAVKMPVGI 273
 +F LI TER+NHARELV L WD + +SGDGL+HEV+NGL++RPDWE A++ P+
 Sbjct: 170 TFKLILTERKNHARELVCAEELGHWDAVMSGDGLMHEVVNGLMERPDWETAIQKPLCS 229

Query: 274 LPCGSGNALAGAVNQHGGEFEPALGLDLLNCSLLLCRGGGHPDLLSVTLASGSRCSFSL 333
 LP GSGNALA +VN + G+E DLL+NC+LLLCR P+LLS+ ASG R +S L
 Sbjct: 230 LPGGSGNALAASVNHAGYEQVTNEDLLINCTLLLCRRRLSPMNLSSLHTASGLRLYSVL 289

Query: 334 SVAWGFVSDVDIQSERFRALGSAREFTLGTVLGLATLHTYGRSLYLPA-TVEPASPTPAH 392
 S++WGFV+DVD++SE++R LG RET+GT LA+L Y+G+L+YLP TV AS PA
 Sbjct: 290 SLSWGFVADVDLESEKYRRLGEIRFTVGTFFRLASLRIYQQLAYLPVGTV--ASKRPAS 347

Query: 393 SL-PRAKSELTLTPDPAPPMH 413
 +L + + L P P +H
 Sbjct: 348 TLVQKGPVDTHLVPLEEPVPSH 369

Score = 324 (48.6 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
 Identities = 72/160 (45%), Positives = 100/160 (62%)

Query: 499 LPLPTPDARVGASTC---GPPDHLLPPLGTPLPDPWVTL-EGDFVLMLAISPSHLGADLV 554
 LP+ T ++ AST GP D L PL P+P W + E DF+L+L + +HL ++L
 Sbjct: 335 LPVGTVASKRPASTLVQKGPVDTHLVPLEEPVPSHWTVPVPEQDFLLVLLHHTLSSELF 394

Query: 555 AAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSFSLGCPQLGYAAARAFLRLEPLT 614
 AAP R + G++HL +VR+G+SRAALLRLFLAM++G H L CP L + AFRLEP +
 Sbjct: 395 AAPMGRCEAGVMHLFYVRAGVSRAALLRLFLAMQKGMELDCPYLVHVPVVAFLRLEPRS 454

Query: 615 PRGVLTVDGEQVEYGPLQAQMHPGIGTLLTGPPGCP-GRE 653
 RGV +VDGE + +Q Q+HP ++ G P GR+
 Sbjct: 455 QRGVFSVDGELMVCEAVQGVHPNYLWMVCGSRDAPSGRD 494

Score = 37 (5.6 bits), Expect = 3.6e-62, Sum P(2) = 3.6e-62
 Identities = 8/20 (40%), Positives = 9/20 (45%)

Query: 459 GAGDAPLSPDPLLSPPGSP 478
 G+ DAP D PP P
 Sbjct: 485 GSRDAPSGRDSRRGPPPEEP 504

Pedant information for DKFZphfbr2_82m6, frame 3

Report for DKFZphfbr2_82m6.3

PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00007	106->115	TYR_PHOSPHO_SITE	PDOC00007
PS00008	56->62	MYRISTYL	PDOC00008
PS00008	212->218	MYRISTYL	PDOC00008
PS00008	232->238	MYRISTYL	PDOC00008
PS00008	272->278	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	279->285	MYRISTYL	PDOC00008
PS00008	361->367	MYRISTYL	PDOC00008
PS00008	476->482	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	574->580	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	640->646	MYRISTYL	PDOC00008
PS00009	122->126	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82m6.3)

DKFZphfkd2_lj9

group: kidney derived

DKFZphfkd2_lj9.3 encodes a novel 105 amino acid protein with high similarity to *Xenopus laevis* XLCL2 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

strong similarity to XLCL2 protein, African clawed frog

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 2955 bp

Poly A stretch at pos. 2935, polyadenylation signal at pos. 2915

```

1  GGGGGGGGCT GAGTGTCTCAG TGGAGAGCGG GGAGTTGTGT CCACCTTGCC
51 GACGTCGCTA GCCGTGGGGC TGTCTGGGA AGGCGGACGG CGAGCGCCCG
101 GTGTCGCGAC TCGGCGCCCT GCCGTGCCCG TCTGCGCCCG TGTCATCCTC
151 ACTCGGGACG CAGGGACCGT TTTTAAATCA CAGGGGCGTG TGTCAGCCTG
201 CCCTAGGACT TCATGTCTAT ATATTTCCCC ATTCACTGCC CCGACTATCT
251 GAGATCGGCC AAGATGACTG AGGTGATGAT GAACACCCAG CCCATGGAGG
301 AGATCGGCCT CAGCCCCCGC AAGGATGGCC TTTCTACCA GATCTTCCCA
351 GACCCGTCAG ATTTTGACCG CCGCTGCAA CTGAAGGACC GTCTGCCCTC
401 CATAGTGGTG GAACCCACAG AAGGGGAGGT GGAGAGCGGG GAGCTCCGGT
451 GGCCCCCTGA GGAGTTCCTG GTCCAGGAGG ATGAGCAAGA TAACTGCCAA
501 GAGACAGCGA AAGAAAAATA AGAGCAGTAG AGTCCCTGTG GACTCCCATG
551 GGTCATACCA GCCAGCATCT GTTCTGAAC TGTGTTTTTC CCATCATGAC
601 GGAAGAAGAG AGTGAGCCGC AATGTGTTCT AAAATGTCAA ACGAGGCTTC
651 TGTTTTGCAC CTGCAGATCA CCGAGTTGGT TTTCTTTTCT TTTCTTGCTT
701 TTTTTTTTTT TTTGAAATTT GCCGAGCAGT GGAGCCCTCT GACAATTTGC
751 AAGGCCCTCT GAGAAAGGAA GCTGCTTAGA GCCAGGGGGT TAGTGGGTGA
801 GGGGAGCGAG TGCTGTTTTT GAGATCATT TCTGAACTCA GGCAGCCTAG
851 TAGAGGCAGT GGTGGGATTC CAATGGGTCT TGGTGGGTGG GAGGTGGGGC
901 ATGTGCAAA CAAGCAAGGA ACATTTGGGG TAAGAAAACA AACATGAGGC
951 AAAAGAAAAA ATACATGTTT TTAAGAAAAC ATTGAGCAGA GAACTGCAGC
1001 CAGGATGCGC TCAGCAGACA TTCACTCTGG CCGCTGGGAC ATCAGAAAAC
1051 AAAGTCTTCA TCTCTCTCTC CAGTTTCACC CACCCACCC TTTGCTTTCA
1101 TTTCAGGTGT GTTGGTCTAT ATGACAGGGA GGAGAGTAAA GGAGAGCAGG
1151 AGCAATTGGC TGCCTGCAAA GCCAGCTGGA GGTGAAGTGC AGGAAAGGAA
1201 AGGTCACCCC ATTCTACTCC ATGGCCTCTC TGCTCCACAG TGTTGGTAGG
1251 TCACATAGCC AGTGTGATCG GTTTTAAAGA GGCAGTGTCT TTCAGCTTTT
1301 CTCCTGATA TATCCATTTT GCTTCCCAGC ACTTTTATAG AGTAGTGAGA
1351 GCACTTCTCT CCCTTGTGTT AAGCCCCAGG GTGGACACTC AGCAGCAAGG
1401 TCTCTCCCTT AACTGCTGCC CTTCCAAGAC TTGCTCCCGA GATGGAGTGG
1451 GCGTGGTCTT CCAGGCTGGC CCTTCTTCTT CCTCACCGCC ACCTTCCCTG
1501 CCCCAGCCCC AGCAGCCATG GGTACATGGG TCCCCAGCTC ACCTATGGAT
1551 TCCCGCCAGT CTGCCAGCT GCAGTACTCA CGCCCCATGG GGGATCTTGG
1601 TCTGTTTTTC TTGTGGGAGC CTAGTGGAGA GCAGACGTGG CTTTTATGTG
1651 GTCTTGTGTT GGAGGTGACT TGCATGGTGG GGACAAGGCT GTCGTGGCAA
1701 CCTTGGGATC GAGTTTGAGA CTAAGGATG TCATGAGATC CCTGGCTTCT
1751 CCCCATGTTT TTCCCGGACA AGGGCAGAAG GGAGGCATGG CAAGGGACCT
1801 CTGCTGTCTT TACTCAACAG TGGTCTCTAT CCCTCCCCAC CTCCCCTGCT
1851 TTCTTGCAAG GGCACCAATT GTATGAGAAA GTTGGCCTTT GGACTTAGGA
1901 TTTCTTATTG TAGCTAAGAG CCATCTGAAG CAGCAGGTGT CAGGACAAAT
1951 GCTTCAGTCC GCCGAGAGCA GTACCGTGTG GCCAAGAGGT GGAATCAGAG
2001 CCTTCTTGA GCTAAACTCG GCCAACCAG GCACGCAGCA TGTCCTCTCA
2051 GGTCTCCAGT CAGTCCAGGT TGACCTCAG TTCTGGACGT GTGTATATAG
2101 CTGTATTTAA TACCTCAAGG TCATTGTGGC TCTGGGGATG CCAGGGCAGG
2151 AGGACGAGGG TCGCTGTGG ACACAGCAGT CCGCGGAATT CCGTCTGGGG
2201 AAGCCAATGG TCGCCGGCAC CCCTTGCTTC CTCCCTCTGT TGTCTGCTGT
2251 TGTGACACAC ATCAATGGCA ATAATTCTT CCAACTCTCT GCAGAAGTGG
2301 GAGAGGCCGG CAGCCTGCAC CGAGAGGGGC TTTCTCTCTT CTGCTCCCC
2351 GCTTCGTTCT GTTTTGGCTG CAGAGAGTGG TTCATCCATA CTCTCATTCC
2401 CTCGCCCTCC CTTGTGGACG GGGGTCTTGC CTTTCAATT CTTGTGTTTT
2451 GGTGTCTTCC CTTATCTGCT ACCCTGAATC ACCTGTCTCT GTCTTGCTGT
2501 GTGATGGGAA CATGCTTGTG AACTGCGTAA CAAATCTACT TTGTGTATGT
2551 GCTGTGTTAT GGGGGTGGTT TATTATTTT GCTGGTCCCT AGACCACTTT
2601 GTATGACCGT TTGCAGTCTG AGCAGGCCAG GGGCTGACAG CTAATGTGAC
2651 GACCTCAGC GGTGGAGCCT GCTGGGGGGA CCCAGCTGCT CTTGGACAAG

```



```

2701 TGGCTGAGCT CCTATCTGGC CTCCTCTTTT TTTTTTTTTT CAAGTAATTT
2751 GTGTGTATTT CTAAGTGATT GTATTGAAAA AATTCCTAGT ATTTAGTAA
2801 AAATGCCTGT TGTGAGATGA ACCTCCTGTA ACTTCTATCT GTTCTTTTTT
2851 GAGGCTCAGG GAGAACTAG CATTTTTTTT TTTCCAAACT ACTTTTGTG
2901 ACTGTGACAG TTGTAAATAA AGTTTGAAAA TGCTCAAAAA AAAAAAAAAA
2951 AAAAC

```

BLAST Results

Entry HSG19750 from database EMBL:
human STS A001X24.
Score = 1050, P = 1.9e-39, identities = 212/213

Entry HSG20267 from database EMBL:
human STS A005C12.
Score = 610, P = 4.1e-19, identities = 122/122

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 213 bp to 527 bp; peptide length: 105
Category: strong similarity to known protein
Classification: unset

```

1 MSIFYPIHCP DYLRSAKMTE VMMNTQPMEE IGLSPRKDGL SYQIFPDPSD
51 FDRCKLKDR LPSIVVEPTE GEVESGELRW PPEEFLVQED EQDNCEETAK
101 ENKEQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_1j9, frame 3

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8e-42

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8.2e-42

>PIR:S52241 XLCL2 protein - African clawed frog
Length = 102

HSPs:

Score = 443 (66.5 bits), Expect = 8.0e-42, P = 8.0e-42
Identities = 80/104 (76%), Positives = 95/104 (91%)

```

Query:      1 MSIFYPIHCPDYLRSAKMTEVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRCKLKDR 60
            MS+++PIHC DYLRSA+MTEV+MNTQ M+EIGLSPRKD SYQIFPDPSDF+R CKLKDR
Sbjct:      1 MSVFYPIHCTDYLRSAEMTEVIMNTQSMDEIGLSPRKD--SYQIFPDPSDFERCCKLKDR 58

Query:      61 LPSIVVEPTEGEVESGELRWPPPEEFLVQEDEQDNCEETAKENKE 104
            LPSIVVEPTEG+VESGELRWPPPEEF+V ED++ C++T KEN++
Sbjct:      59 LPSIVVEPTEGDVESGELRWPPPEEFVVDKEDKGTCDQTKKENEQ 102

```

Pedant information for DKFZphfkd2_1j9, frame 3

Report for DKFZphfkd2_1j9.3

```

[LENGTH]      105
[MW]           12269.78
[pI]           4.40
[HOMOL]        PIR:S52241 XLCL2 protein - African clawed frog 5e-44

```


{KW} Alpha_Beta

SEQ MSYFPIHCPDYLRSAKMTVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRCKLKDR
PRD cccccccccchhhhhhhhhhhccccccccccccccccccccccccchhhhhhhc

SEQ LPSIVVEPTEGEVESGELRWPPEEFLVQEDEQDNCEETAKENKEQ
PRD cccceeeccccccccccccccccccccccccchhhhhhhhhccc

(No Prosite data available for DKFZphfd2_1j9.3)

(No Pfam data available for DKFZphfd2_1j9.3)

DKFZphfkd2_24a15

group: transmembrane protein

DKFZphfkd2_24a15 encodes a novel amino acid protein with similarity to C. elegans cosmid R07G3.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to C. elegans R07G3.8
membrane regions: 1
Summary DKFZphfkd2_24a15 encodes a novel 323 amino acid protein, with
similarity to C. elegans R07G3.8.

similarity to C. elegans R07G3.8

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1513 bp
Poly A stretch at pos. 1494, no polyadenylation signal found

```
1 GGGGTACTCG GCGGCGGCGG AGCGGGGCGG AGAGCAGGGC GCGGCGGACT
51 CGCAGGGTAC CACCATCTTA AGGACAGAAA AGCTACAGGA CTCTAGGAGG
101 CCACCGTCCT GATTGGGAA GTCCAACCTTA CTTGGCCAG ACAGCAGCTA
151 AGCTGGTTCA TCCCATCAGC CTGGATTGGT GAAACTGAAT CACAGGAGAT
201 ATTTCCAGGT TTGCTGGGAT GGGAAACCTG CTCAAAGTCC TTACCAGGGA
251 AATTGAAAAC TATCCACACT TTTTCCTGGA TTTTGAAAAT GCTCAGCCTA
301 CAGAAGGAGA GAGAGAAATC TGGAACCAGA TCAGCGCCGT CCTTCAGGAT
351 TCTGAGAGCA TCCTTGCGA CCTGCAGGCT TACAAAGCCG CAGGCCGAGA
401 GATCCGAGAT GCAATTCAAA ATCCCAATGA CATTGAGCTT CAAGAAAAAG
451 CTTGGAATGC GGTGTGCCCT CTTGTTGTGA GGCTAAAGAG ATTTTACGAG
501 TTTTCCATTA GACTAGAAAA AGCTCTTCAG AGTTTATTGG AATCTCTGAC
551 TTGTCCACCC TACACACCAA CCAACACCTT GGAAGGGGAA CAGGCCCTGG
601 CAAAGGAGTT TGCCGAAATT TTACATTTTA CCCTTCGATT CGATGAGCTG
651 AAGATGAGGA ACCCGGCTAT TCAGAATGAC TTCAGCTACT ACAGAAGAAC
701 AATCAGTCGC AACCGCATCA ACAACATGCA CCTAGACATT GAGAATGAAG
751 TCAATAATGA GATGGCCAAT CGAATGTCCC TCTCTATGC AGAAGCCACG
801 CCAATGCTGA AAACCCCTAG CAATGCCACA ATGCACCTTG TCTCTGAAAA
851 CAAACTCTG CCAATAGAGA ACACCACAGA CTGCCTCAGC ACAATGACAA
901 GTGTCTGTAA AGTCATGCTG GAAACTCCG AGTACAGAAG TAGGTTTACG
951 AGTGAAGAGA CCCTGATGTT CTGCATGAGG GTGATGGTGG GAGTCATCAT
1001 CCTCTATGAC CATGTCCACC CTGTGGGAGC TTTCTGCAAG ACATCCAAGA
1051 TCGATATGAA AGGCTGCATA AAAGTTTGA AGGAGCAGGC CCCAGACAGT
1101 GTGGAGGGGC TGCTAAATGC CCTCAGGTTT ACTACAAAGC ACTTGAACGA
1151 TGAATCAACT TCCAAACAGA TTCGAGCAAT GCTTCAGTAG AGCTCTGCTC
1201 AAAGAAGAGG ATCTATGTGC TGACCTCAGA AGATGTATAT GTTTACATAA
1251 TTTAATACAG ATTGATGTTA ATACTTGTGT ATTTACATAA CCGTTTCCTT
1301 CTGTCACTG AAATATATGG ACCTTAATTT GTATCCTGAC TGAATCAACC
1351 CAGCAGAGCA TAAATTGACT TGAGAGCCTT ACCTTTGATG TCTGAAATGA
1401 AACCCCTTTC TCCAAAGGCA AAATTCGGAG ACTTTGATCT TTGCTACTGG
1451 AGTCCTTTAA CAACATCTAT AACGATAAAA AATTCCTAAT TGTCAAAAAA
1501 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 219 bp to 1187 bp; peptide length: 323
 Category: similarity to unknown protein

```

1 MGNNLLKVLTR EIENYPHFLL DFENAOPTTEG EREIWNQISA VLQDSESILA
51 DLQAYKGAGP EIRDAIQNP DIQLQEKAWN AVCPLVVRK RFYEFSIRLE
101 KALQSLLES L TCPPTPTQH LEREQALAKE FAEILHFTLR FDELKMRNPA
151 IQNDFSYYRR TISRNRINNM HLDIENEVNN EMANRMSLFY AEATPMLKTL
201 SNATMHFVSE NKTLPIENTT DCLSTMTSVC KVMELETPEYR SRFTSEETLM
251 FCMRVMVGVI ILYDHVHPVG AFCKTSKIDM KGCIVLKEQ APDSVEGLLN
301 ALRFTTKHLN DESTSKQIRA MLQ

```

BLASTP hits

Entry CER07G3_7 from database TREMBL:
 gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3.
 Score = 544, P = 1.4e-52, identities = 119/323, positives = 186/323

Alert BLASTP hits for DKFZphfd2_24a15, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfd2_24a15, frame 3

Report for DKFZphfd2_24a15.3

```

[LENGTH] 323
[MW] 37313.06
[pI] 5.71
[HOMOL] TREMBL:CER07G3_7 gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3. 4e-54

[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 5
[PROSITE] ASN_GLYCOSYLATION 3
[KW] TRANSMEMBRANE 1

```

```

SEQ MGNNLLKVLTR E IENYPHFLL DFENAOPTTEG EREIWNQISA VLQDSESILA DLQAYKGAGP
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ EIRDAIQNPNDIQLQEKAWN AVCPVVRK RFYEFSIRLE KALQSLLES L TCPPTPTQH
PRD hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ LEREQALAKEFAEILHFTLR FDELKMRNPA IQNDFSYYRR TISRNRINNM HLDIENEVNN
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ EMANRMSLFYAEATPMLKTL SNATMHFVSE NKTLPIENTT DCLSTMTSVC KVMELETPEYR
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ SRFTSEETLMFCMRVMVGVI ILYDHVHPVG AFCKTSKIDM KGCIVLKEQ APDSVEGLLN
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ ALRFTTKHLN DESTSKQIRA MLQ
PRD hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

```

Prosite for DKFZphfd2_24a15.3

PS00001	202->206	ASN_GLYCOSYLATION	PDOC00001
PS00001	211->215	ASN_GLYCOSYLATION	PDOC00001
PS00001	218->222	ASN_GLYCOSYLATION	PDOC00001
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	275->278	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005

PS00005	314->317	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00007	231->240	TYR_PHOSPHO_SITE	PDOC00007
PS00008	297->303	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_24a15.3)

DKFZphfkd2_24b15

group: metabolism

DKFZphfkd2_24b15 encodes a novel 612 amino acid protein with similarity to bacterial and yeast phosphoglucosyltransferase and phosphomannosyltransferases.

The novel protein contains a phosphoserine signature typical for phosphoglucosyltransferase (EC 5.4.2.2) or phosphomannosyltransferase (EC 5.4.2.8). Thus, the protein seems to be taking part in the conversion of hexose phosphates.

The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

similarity to phosphomannosyltransferases

complete cDNA, complete cds, EST hits
potential start at bp 30 matches kozak consensus PyCnatgG,

Sequenced by GBF

Locus: map="158.8 cR from top of Chr4 linkage group"

Insert length: 2204 bp
Poly A stretch at pos. 2186, no polyadenylation signal found

```
1 GGGCTCTGCA GCGGTAGCAC AAGCTCAGCG ATGGCGGCTC CAGAAGGCAG
51 CGGTCTAGGC GAGGACGCCG GGCTGGACCA GGAGACGCCG CAGTGGCTGC
101 GCTGGGACAA GAATTCCTTA ACTTTGGAGG CAGTGAACGC ACTAATAGCA
151 GAAGGTAATA AAGAAGAACT ACGAAAATGT TTTGGGGCCC GAATGGAGTT
201 TGGGACAGCT GGCCTCCGAG CTGCTATGGG ACCTGGAATT TCTCGTATGA
251 ATGACTTGAC CATCATCCAG ACTACACAGG GATTTTGCAG ATACCTGGAA
301 AAACAATTCA GTGACTTAAA GCAGAAAGGC ATCGTGATCA GTTTTGACGC
351 CCGAGCTCAT CCATCCAGTG GGGGTAGCAG CAGAAGGTTT GCCCGACTTG
401 CTGCAACCAC ATTTATCAGT CAGGGGATTC CTGTGTACCT CTTTCTGAT
451 ATAACGCCAA CCCCCTTTGT GCCCTTCACA GTATCACATT TGAAACTTTG
501 TGCTGGAATC ATGATAACTG CATCTCACAA TCCAAAGCAG GATAATGGTT
551 ATAAGGTCTA TTGGGATAAT GGAGCTCAGA TCATTCTTCC TCACGATAAA
601 GGGATTTTCT AAGCTATTGA AGAAAATCTA GAACCGTGGC CTCAAGCTTG
651 GGACGATTCT TTAATTGATA GCACTCCACT TCTCCACAAT CCGAGTGCTT
701 CTCCTGAGGC TGACTACTTT GAAGACCTTA AAAAGTACTG TTTCCACAGG
751 AGCGTGAACA GGGAGACAAA GGTGAAGTTT GTGCACACCT CTGTCCATGG
801 GGTGGGTGAT AGCTTTGTGC AGTCAGCTTT CAAGGCTTTT GACCTTGTTT
851 CTCCTGAGGC TGTTCCTGAA CAGAGAGATC CGGATCCTGA GTTTCCAACA
901 GTGAAATACC CGAATCCCGA AGAGGGGAAA GGTGTCTTGA CTTTGTCTTT
951 TGCTTTGGCT GACAAAACCA AGGCCAGAAT TGTTTAGCT AACGACCCGG
1001 ATGCTGATAG ACTTGCTGTG GCAGAAAAGC AAGACAGTGG TGAATGGAGG
1051 GTGTTTTTCA GCAATGAGTT GGGGGCCCTC CTGGGCTGGT GGCTTTTAC
1101 ATCTTGGAAG GAGAAGAACC AGGATCGCAG TGCTCTCAAA GACACGTACA
1151 TGTTGTCCAG CACCGTCTCC TCCAAAATCT TCGGGGCCAT TGCCTTAAAG
1201 GAAGGTTTTT ATTTTGAGGA AACATTAACCT GGCTTTAAGT GGATGGGAAA
1251 CAGAGCCAAA CAGCTAATAG ACCAGGGGAA AACTGTTTTA TTTGCATTG
1301 AAGAAGCTAT TGGATACATG TGCTGCCCTT TTGTTCTGGA CAAAGATGGA
1351 GTCAGTGCCG CTGTCATAAG TGCAGAGTTG GCTAGCTTCC TAGCAACCAA
1401 GAATTTGTCT TTGTCTCAGC AACTAAAGGC CATTATGTG GAGTATGGCT
1451 ACCATATTAC TAAAGCTTCC TATTTTATCT GCCATGATCA AGAAACCATT
1501 AAGAAATTAT TTGAAAACCT CAGAACTAC GATGGAAAAA ATAATTATCC
1551 AAAAGCTTGT GGCAAATTTG AAATTTCTGC CATTAGGGAC CTTACAACCTG
1601 GCTATGATGA TAGCCAACCT GATAAAAAAG CTGTTCTTCC CACTAGTAAA
1651 AGCAGCCAAA TGATCACCTT CACCTTTGCT AATGGAGGCG TGGCCACCAT
1701 GCGCACCAGT GGGACAGAGC CCAAAATCAA GTACTATGCA GAGCTGTGTG
1751 CCCCACCTGG GAACAGTGAT CCTGAGCAGC TGAAGAAGGA ACTGAATGAA
1801 CTGGTCAGTG CTATTGAAGA ACATTTTTTC CAGCCACAGA AGTACAATCT
1851 GCAGCCAAAA GCAGACTAAA ATAGTCCAGC CTTGGGTATA CTTGCATTTA
1901 CCTACAATTA AGCTGGGTTT AACTTGTTAA GCAATATTTT TAAGGGCCAA
1951 ATGATTCAAA ACATCACAGG TATTTATGTG TTTTACAAAG ACCTACATTC
2001 CTCATTGTTT CATGTTTGAC CTTTAAGGTG AAAAAAGAAA ATGGCCAAAC
2051 CCAACAAACT AACATTCCTA CTAAAAAGTT GAGCTTGGAC ATATTTTGAA
2101 TTTTGTGAAG TGAAGATTTT TAACTGACT AACTTAAAAA AATAGATTGT
2151 AATTGATGTG CCTTAATTG CATAAATCAT AAATGTAAAA AAAAAAAAAA
2201 AAAA
```

BLAST Results

Entry HS705145 from database EMBL:

human STS WI-6820.

Score = 1261, P = 3.6e-52, identities = 253/254

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 31 bp to 1866 bp; peptide length: 612
 Category: strong similarity to known protein

```

1 MAAPEGSGLG EDARLDQETA QWLRWDKNSL TLEAVKRLIA EGNKEELRKC
51 FGARMEFGTA GLRAAMGPGI SRMNDLTIIQ TTQGFRCYLE KQFSDLKQKG
101 IVISFDARAH PSSGGSSRRF ARLAATTFIS QGIPVYLFSD ITPTFPVPFT
151 VSHLKLKAGI MITASHNPKQ DNGYKVYWDN GAQIISPHDK GISQAIEENL
201 EPWPQAWDDS LIDSSPLLHN PSASINNDYF EDLKKYCFHR SVNRETKVKE
251 VHTSVHGVGH SFVQSFAKAF DLVPPPEAVPE QRDPEFPFT VKYPNPEEGK
301 GVLTLSEFALA DTKARIVLA NDPDADRLAV AEKQDSGEWR VFSGNELGAL
351 LGWWLFTSWK EKNQDRSALK DTYMLSSTVS SKILRAIALK EGFHFEETLT
401 GFKWMGNRAK QLIDQGTIVL FAFEEAIGYM CCPFVLDKDG VSAAVISAEI
451 ASFLATKNLS LSQOLKAIYV EYGYHITKAS YFICHQDQETI KKLFEENLRNY
501 DGKNNYPKAC GKFEISAIRD LTTGYDDSQP DKKAVLPTSK SSQMITFTFA
551 NGGVATMRTS GTEPKIKYYA ELCAPPNGSD PEQLKKELNE LVSAIEEHFF
601 QPQKYNLQPK AD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfkd2_24b15, frame 1

TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B, N = 1, Score = 1431, P = 1.6e-146

TREMBL:SPCC1840_5 gene: "SPCC1840.05c"; product: "similarity to phosphomannomutases"; S.pombe chromosome III cosmid c1840., N = 1, Score = 1210, P = 4.2e-123

PIR:S54585 hypothetical protein YMR278w - yeast (Saccharomyces cerevisiae), N = 1, Score = 1046, P = 1e-105

PIR:A71299 probable phosphomannomutase (manB) - syphilis spirochete, N = 1, Score = 697, P = 9.7e-69

>TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B Length = 595

HSPs:

Score = 1431 (214.7 bits), Expect = 1.6e-146, P = 1.6e-146
 Identities = 285/598 (47%), Positives = 393/598 (65%)

```

Query:   13 ARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKC FGARMEFGTAGLRAAMGPGISR 72
          A+LD++ A WL WDKN      +++L+ E N + L+      R+ FG TAG+R+ M G R
Sbjct:   6 AKLDKQVADWLAWDKNDKNRNEIQKLVDKENVDAKARMDTRLVFGTAGVRSMPQAGFGR 65

Query:   73 MNDLTIIQTTQGFRCYLEKQFSDLKQKGIVISFDARAH PSSGGSSRRFARLAATTFISQG 132
          +NDLTIIQ T GF R++      + K G+ I FD R +      SRRFA L+A F+
Sbjct:   66 LNDLTIIQITHGFARHMLNVYGPKN-GVAIGFDGRYN-----SRRFAELSANVFVRNN 118

Query:   133 IPVYLFSDITPTFPFPFTVSHLKLKAGIMITASHNPKQDNGYKVYWDNGAQIISPHDKGI 192
          IPVYLFSD+++PTP V +      L AG++ITASHNPK+DNGYK YW NGAQII PHD I
Sbjct:   119 IPVYLFSEVSPFPVVSATIKLGCDAGLIITASHNPKEDNGYKAYWSNGAQIIGPHDTEI 178

Query:   193 SQAIEENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHRSVNRETKVKEVH 252
          + E +P + WD S + SSPL H+      I+ YFE K F R +N T +KF +
Sbjct:   179 VRIKEAEPQPRDEYWDLSLKSSPLFHSADVVID-PYFEVEKSLNFTREINGSTPLKFTY 237

Query:   253 TSVHGVGHGSFVQSFAKAFDLVPPPE--AVPEQRDPDPEFPFTVKYPNPEEGKGVLTLSFALA 310
          ++ HG+G+ + + F F      +V EQ+DP+P+PFT+ +PNPEEG+ VLT L+ A
Sbjct:   238 SAFHGIGYHYTKRMFAEFGFPPASSFISVAEQQDPNPDPFTPIFPNPEEGKRVLTLMETA 297

```


Query: 311 DTKKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWLFWSWKEKNQDRSALK 370
 DK + ++LANDPDADR+ +AEKQ GEWRVF+GNE+GAL+ WW++T+W++ N + A K
 Sbjct: 298 DKNGSTVILANDPDADRIQMAEKQKDGGEWRVFTGNEMGALITWWIWTNWRKANPNADASK 357

Query: 371 DTYMLSSSTVSSKILRAIALKEGFHEETLTGFKWMGNRAKQLIDQKTVLFAFEEAIGYM 430
 Y+L+S VSS+I++ IA EGF E TLTGFKWMGNRA++L G V+ A+EE+IGYM
 Sbjct: 358 -VYILNSAVSSQIVKTIADAEGFKNETLTGFKWMGNRAEELRADGNQVILAWEEESIGYM 416

Query: 431 CCP-FVLDDKGVSAAVISAEASFLATKNLSLSQLKAIYVEYGYHITKASYFICHQDET 489
 P +DKDGVSA + AE+A+FL + SL QL A+Y YG+H+ +++Y++ E
 Sbjct: 417 --PGHTMDKGVSAAAVFAEIAAFLHAEGKSLQDQLYALYNRYGFHLVRSTYWMVPAPEV 474

Query: 490 IKKLFENLRNYDGKNYPKACGKFEISAIRDLTTGYDDSQPKKAVLPTSKSSQMITFTF 549
 KKLF LR D K +P G+ E++++RDLT GYD+S+PD K VLP S SS+M+TF
 Sbjct: 475 TKKLFSTLRA-DLK--FPTKIGAEVASVRDLTIGYDNSKPDNKPVLPLSTSEMVTFFL 531

Query: 550 ANGGVATMRTSGTEPKIKYYAELCAPPGNS--DPEQLKKELNELVSAIEEHFFQPKYKYL 607
 G V T+R SGTEPKIKYY EL PG + D E + E+++L + +PQ++ L
 Sbjct: 532 KTGSVTTLRASGTEPKIKYYIELITAPGKTQNDLESVISEMDQLEKDVVATLLRPQQFGL 591

Query: 608 QPK 610
 P+
 Sbjct: 592 IPR 594

Pedant information for DKFZphfkd2_24b15, frame 1

Report for DKFZphfkd2_24b15.1

[LENGTH] 612
 [MW] 68311.58
 [pI] 6.28
 [HOMOL] TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B le-157

[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YMR278w] 1e-111
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0740] 3e-66
 [FUNCAT] c energy conversion [M. genitalium, MG053] 4e-50
 [FUNCAT] m outer membrane and cell wall [H. influenzae, HI1463] 2e-04
 [BLOCKS] BL00607D cAMP phosphodiesterases class-II proteins
 [BLOCKS] BL00710 Phosphoglucosyltransferase and phosphomannomutase phosphoserine signa
 [EC] 5.4.2.8 Phosphomannomutase 3e-56
 [EC] 5.4.2.2 Phosphoglucosyltransferase 1e-09
 [PIRKW] isomerase 3e-56
 [PIRKW] intramolecular transferase 3e-56
 [SUPFAM] Methanobacterium thermoautotrophicum phosphomannomutase 1e-06
 [SUPFAM] probable phosphorylating protein ureC 9e-06
 [PROSITE] PGM_PMM1
 [PROSITE] MYRISTYL 10
 [PROSITE] LIPOCALIN 2
 [PROSITE] CK2_PHOSPHO_SITE 9
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Phosphoglucosyltransferase and phosphomannomutase phosphoserine
 [KW] Alpha_Beta

SEQ MAAPEGSGLGEDARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKCFCGARMFEGTA
 PRD ccc

SEQ GLRAAMGPGISRMNDLTIIQTQGFRCRYLEKQFSDLKQKGIVISFDARAHPPSSGGSSRRF
 PRD ccc

SEQ ARLAATTFISQIGIPVYLFSDITPTFPVFTVSHLKLKAGIMITASHNPKQDNGYKVYWDN
 PRD hhhhhhhhhhhcc

SEQ GAQIISPHDKGISQAIENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHR
 PRD ccc

SEQ SVNRETKVKFVHTSVHGVGHSFVQSAFAKFDLVPPEAVPEQRDPDFEPTVKYPNPEEGK
 PRD ccc

SEQ GVLTLSFALADTKKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWLFWSW
 PRD hhhhhhhhhhhhhcc

SEQ EKNQDRSALKDQTYMLSSSTVSSKILRAIALKEGFHEETLTGFKWMGNRAKQLIDQKTVL
 PRD hcc


```

SEQ  FAFEEAIGYMCCPFVLDKDGVSAAVISAE LASFLATKNLSLSQQLKAIYVEYGYHITKAS
PRD  hhhhhccccccccccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhccccccccc

SEQ  YFICHQDETIKKLFENLRNYDGKNYPKACGKFEISAIRDLTTGYDDSQPKKAVLPTSK
PRD  eeecchhhhhhhhhhhhhhhhhccccccccccccchhhhhhhccccccccccccccccccccccccccc

SEQ  SSQMITFTTFANGGVATMRTSGTEPKIKYYAELCAPPGNSDPEQLKKELNELVSAIEEHFF
PRD  cccceeeeeccccceeeccccccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  QPQKYNLQPKAD
PRD  cccccccccccc

```

Prosites for DKFZphfkd2_24b15.1

PS00001	458->462	ASN_GLYCOSYLATION	PDOC00001
PS00002	7->11	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	380->383	PKC_PHOSPHO_SITE	PDOC00005
PS00005	489->492	PKC_PHOSPHO_SITE	PDOC00005
PS00005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS00005	556->559	PKC_PHOSPHO_SITE	PDOC00005
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	528->532	CK2_PHOSPHO_SITE	PDOC00006
PS00006	560->564	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	61->67	MYRISTYL	PDOC00008
PS00008	100->106	MYRISTYL	PDOC00008
PS00008	159->165	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	257->263	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00008	348->354	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00710	159->174	PGM_PMM	PDOC00589
PS00213	346->358	LIPOCALIN	PDOC00187
PS00213	344->358	LIPOCALIN	PDOC00187

Pfam for DKFZphfkd2_24b15.1

```

HMM_NAME      Phosphoglucomutase and phosphomannomutase phosphoserine
HMM            *GvnVidIGQNGMMPTPMIYFaIRTYKhmcmggGIMITaSHNPGGPDnDN
               G+ V + ++PTP + F + H+++ +GIMITASHNP DN
Query         132 GIPVYLFS--DITPTFPVPFTVS---HLKLCAGIMITASHNP--KQ-DN 172
HMM            GIK*
               G+K
Query         173 GYK 175

```


DKFZphfkd2_24e23

group: kidney derived

DKFZphfkd2_24e23 encodes a novel 198 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit,
many ATGs in front of the ORF

Sequenced by GBF

Locus: unknown

Insert length: 1723 bp

Poly A stretch at pos. 1695, no polyadenylation signal found

```
1 GGGGGATTTC CGATCATGAC AACGATAGCA ATTGATATAC CTTCAAAATA
51 CGTGTCCAGT GAGTGTGGT TGTGTGGT TTCTCTAGGA GACCGTGTTC
101 ATGCAACACA GCATTATTTC ACCGCCITTA CCCCAGCTTC TTCATACACA
151 TGCACCTTGT CAGGGCTCTT TGGCTGAAGA GAAGTTAGAA GTTCCAGAT
201 ATGGAGGGGT ATTTTCAGCA GATATGCCCA CCGCCATGGT TTTGTCAGCT
251 CTGTAGGGTG GTCTTGACCC CTGCTCACTG CTGGCATCAC CTGAGCCTAT
301 GGCAGATACC CAGTGCTGCC CGCCACCATG TGAATTCATC AGCTCTGCAG
351 GCACAGACCT TGCAC TAGGA ATGGGCTGGG ACGCCACCCT CTGCCTCTTA
401 CCATTCACTG GGTTTGGCAA GTGTGCTGGG ATCTGGAATC ACATGGATGA
451 GGAACCCGAT AATGGTGACG ACCGAGGTAG CAGGCGAACC ACTGGCCAGG
501 GCAGGAAGTG GGCAGCTCAC GGGACTATGG CTGCACCGCG GGTTCATACC
551 GACTACCATC CTGGAGGTGG GAGCGCATGC TCATCTGTAA AAGTCCGGTC
601 CCACGTTGGA CACACCGGGG TCTTCTCTT TGTGACCAG GATCCTCTGG
651 CAGTGTCTTT AACAAAGCCAG AGTCTGATCC CACCGCTCAT AAAGCCAGGG
701 TTGTTGAAAG CTTGGGGCTT CCTCCTCCTC TGTGCGCAGC CCTCAGCAAA
751 CGGTCACAGC CTGTGCTGTC TGCTGTACAC CGACTTGGTA TCATCCCATG
801 AACTGTCCCC CTTTCGTGCT CTGTGCTTAG GGGCCTCTGA TGCCCCATCT
851 GCCTGCGCTT CCTGCAACTG TTTAGCAAGC ACCTATTATC TATAGGGTGC
901 TGGGGTGCTG GGCAGAGCCA ATCGCTCCTA TTACTTTCTG CCCTGGGGAC
951 GTCCTGTTTT CCCACCTACC CCTGTAAACG CTCTGCTCTG CCTTCCCATC
1001 TGCGGGGCTA ACGCCATCCC ACAAGGGCTG GGCTGTCCGT TCAGAAGAGA
1051 AACTGGGAAG GGGCCTTGAG GACCTGTGTC CAGGACGGGT GGACAAGGGC
1101 TTTGTGCAGG GAGCTCCTCT CCCATCTTTG TGTCTGACA GCCGTGACCG
1151 TGACCCCTCA AAGCAGAGCC AGTAGTGATC AGTATCCTGC TGCTTCAAGC
1201 CTGCACGGTC CTCTTCTCCT CTCCGCACAT CTGCATGCCT GTCAAACCCA
1251 GAGTAGTTTG GGGCCTGGTA AACAGAGGGA AGTTGGCTGG AGGAGGCCAG
1301 TCAGGAGTGC AAGAACCCCG CGTACTCTGT CCCACGTGGA TAAAGTCTCT
1351 AATTCCAGTC TGAGGTGAAT TCTTAGAGAG TGCTTTCATT TAATGTTTGC
1401 TTTATGCATT TCCCCTGCAG CTGTGACTAA TTGTGGAACA GCATACATTT
1451 TGTTTTGAGA CTCTCTTGAG ATTTTCTGG CAGTGTAAGG TCTACACCAT
1501 TTTCTCTCA GCATCAGAGA AGGCAGAAAG CAAGAGAAAG GAATGCAATG
1551 TGAGCAAGGC CAGGCACACT TGTGCTACTG CAGTTGGCAA GAATGGAGTC
1601 TAATCCAGC ACTTTGGGAG GCCGAGGCGG GTGGATCACC TGAGGTCAAG
1651 AATTTGAGAC CAACCTGGCC AACATGTTGA AACCTCGTCT GACTAAAAA
1701 TACAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 299 bp to 892 bp; peptide length: 198
Category: putative protein

```

1 MADTQCCPPP CEFISSAGTD LALGMGWDAT LCLLPFTGFG KCAGIWNHMD
51 EEPDNGDDRG SRRTTGQGRK WAAHGTMAAP RVHTDYHPGG GSACSSVKVR
101 SHVGHTGVFF FVDQDPLAVS LTSQSLIPPL IKPGLLKAWG FLLCAQPSA
151 NGHSLCCLLY TDLVSSHELSPFRALCLGPS DAPSACASCN CLASTYYL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24e23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_24e23, frame 2

Report for DKFZphfkd2_24e23.2

```

[LENGTH]      198
[MW]           20948.98
[PI]           6.01
[PROSITE]      MYRISTYL      5
[PROSITE]      AMIDATION     1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 2
[KW]           All_Beta
[KW]           LOW_COMPLEXITY 6.06 %

```

```

SEQ  MADTQCCPPPCFEISSAGTDLALGMGWDATLCLLPFTGFGKCAGIWNHMDDEPDNGDDRG
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SRRTTGQGRKWAHGTMAAPRVHTDYHPGGGSACSSVKVRSHVGHTGVFFVDQDPLAVS
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LTSQSLIPPLIKPGLLKAWGFLLLCAQPSANGHSLCCLLYTDLVSSHELSPFRALCLGPS
SEG  .....XXXXXXXXXXXX.....
PRD  eccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ  DAPSACASCNCLASTYYL
SEG  .....
PRD  ccccccccccccccccccc

```

Prosites for DKFZphfkd2_24e23.2

PS00004	62->66	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	60->66	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008
PS00008	91->97	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfkd2_24e23.2)

DKFZphfkd2_24n20.

group: intracellular transport and trafficking

DKFZphfkd2_24n20.3 encodes a novel 366 amino acid protein with similarity to human eps8 binding protein e3B1 and spectrins.

The new protein contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. Eps8 is a substrate of receptor tyrosine kinases involved in mitogenic signaling. Spectrin is part of the submembrane cytoskeletal network in the human erythrocyte ghost. Nonerythroid spectrins are proposed to have roles in cell adhesion, establishment of cell polarity, and attachment of other cytoskeletal structures to the plasma membrane. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

strong similarity to eps8 binding protein e3B1

complete cDNA, complete cds, few EST hits
potential start at Bp 300, but there are ATGs in other frames in
5' region of the cDNA

Sequenced by GBF

Locus: /map="17"

Insert length: 1719 bp
Poly A stretch at pos. 1699, polyadenylation signal at pos. 1680

```

1 GGGGACAGCT GCCCCGACCT TGGCTTCCTC TGCTGGGTGG GATTGGGGGC
51 TGGGGCCCCA AATGGGGCCC TGGCTTCCCC CTTCCTCTGG GCAGGGGACA
101 GAGAGACACA GGCTCGGGGA GCAGGACTGA CTTCTCTTGG TCCCGGAATG
151 AGCATGCCTG CCCTTTGCAA GCAGGTTTGG GTCTCACGCA GAGGAAACCA
201 AAAGCAATAA GAGGGAGGGA AGGCAGAGCA ACCAATCAAG GGCAGGGTGA
251 GACTCAAAAC GAGCGGGCTC CCTGGGGAGC CAGACAGAGG CTGGGGGTGA
301 TGGCGGAGCT ACAGCAGCTG CAGGAGTTTG AGATCCCCAC TGGCCGGGAG
351 GCTCTGAGGG GCAACCACAG TGCCCTGCTG CGGGTCGCTG ACTACTGCGA
401 GGACAACATAT GTGCAGGCCA CAGACAAGCA GAAGGCGCTG GAGGAGACCA
451 TGGCCTTCAC TACCCAGGCA CTGGCCAGCG TGGCCTACCA GGTGGGCAAC
501 CTGGCCGGGC ACACCTGCGG CATGTTGGAC CTGCAGGGGG CCGCCCTGCG
551 GCAGGTGGAA GCCCGTGTA GCACGCTGGG CCAGATGGTG AACATGCATA
601 TGGAGAAGGT GGCCCGAAGG GAGATCGGCA CCTTAGCCAC TGTCCAGCGG
651 CTGCCCCCGG GCCAGAAGGT CATCGCCCCA GAGAACCTAC CCCCTCTCAC
701 GCCCTACTGC AGGAGACCCC TCAACTTTGG CTGCCTGGAC GACATTGGCC
751 ATGGGATCAA GGACCTCAGC ACGCAGCTGT CAAGAACAGG CACCCTGTCT
801 CGAAAGAGCA TCAAGGCCCC TGCCACACCC GCCTCCGCCA CCTTGGGGAG
851 ACCGCCCCGG ATTCCCGAGC CAGTGCACCT GCCGGTGGTG CCCGACGGCA
901 GACTCTCCCG CGCCTCCTCT GCGTCTTCCC TGGCCTCGGC CGGCAGCGCC
951 GAAGGTGTCT GTGGGGCCCC CACGCCAAGG GGGCAGGCAG CACCTCCAGC
1001 CCCACCTCTC CCCAGCTCCT TGGACCCACC TCCTCCACCA GCAGCCCTCG
1051 AGGTGTTCCA GCGGCCTCCC ACGCTGGAGG AGTTGTCCCC ACCCCACCG
1101 GACGAAGAGC TGCCCTTGCC ACTGGACCTG CCTCCTCCTC CACCCTTGGA
1151 TGGAGATGAA TTGGGGCTGC CTCCACCCCC ACCAGGATTT GGGCCTGATG
1201 AGCCCAAGTG GGTGCCTGCC TCATACTTGG AGAAAGTGGT GACACTGTAC
1251 CCATACACCA GCCAGAAGGA CAATGAGCTC TCCTTCTCTG AGGGCACTGT
1301 CATCTGTGTC ACTCGCCGCT ACTCCGATGG CTGGTGCGAG GGCCTCAGCT
1351 CGGAGGGGAC TGGATTCTTC CCTGGGAAC TGTGGAGGCC CAGCTGCTGA
1401 CAGCCCAGGG CTCTCTGGGC AGCTGATGTC TGCACTGAGT GGGTTTCATG
1451 AGCCCCAAGC CAAAACCAGC TCCAGTCACA GCTGGACTGG GTCTGCCCCA
1501 CTCTTTGGCT GTGAGCTGTG TTCTGTCTTT CCTCCCATCG GAGGGAGAAG
1551 GGGTCTCTGG GAGAGAGAAT TTATCCAGAG GCCTGCTGCA GATGGGGAAG
1601 AGCTGGAAAC CAAGAAGTTT GTCAACAGAG GACCCCTACT CCATGCAGGA
1651 CAGGGTCTCC TGCTGCAAGT CCCAACTTTG AATAAACAG ATGATGTCCA
1701 AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry AC004797 from database EMBL:
Homo sapiens chromosome 17, clone hRPC.62 O.9, complete sequence.
Score = 2316, P = 5.9e-255, identities = 464/465
7 exons Bp 93317-110902

Medline entries

97163405:
Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth.

98256293:
Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton.

Peptide information for frame 3

ORF from 300 bp to 1397 bp; peptide length: 366
Category: strong similarity to known protein

```

1 MAELQQLQEF EIPTGREALR GNHSALLRVA DYCEDNYVQA TDKQKALEET
51 MAFTTQALAS VAYQVGNLAG HTLRMLDLQG AALRQVEARV STLGMVMNMH
101 MEKVARREIG TLATVQRLPP GQKVIAPENL PPLTPYCRRP LNFGCLDDIG
151 HGIKDLSTQL SRTGTLRKS IKAPATPASA TLGRPPRIPE PVHLPVVPDG
201 RLSAASSASS LASAGSAEGV GGAPTPKQQA APPAPPLPSS LDPPPPPAAV
251 EVFQRPPTLE ELSPPPPDEE LPLPLDLPPP PPLDGLGL PPPPPGFGPD
301 EPSWVPASYL EKVVTLYPYT SQKDNELSFS EGTVICVTRR YSDGWCEGVS
351 SEGTFGFFGN YVEPSC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfk2_24n20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfk2_24n20, frame 3

Report for DKFZphfk2_24n20.3

```

[LENGTH]      366
[MW]           38947.21
[pI]           4.93
[HOMOL]        TREMBL:U87166 1 gene: "SSH3BP1"; product: "spectrin SH3 domain binding protein
1"; Homo sapiens spectrin SH3 domain binding protein 1 (SSH3BP1) mRNA, complete cds. 3e-48

[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YPR154w] 3e-05
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YDR388w] 2e-04
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR388w]
2e-04
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDR162c] 4e-04
[BLOCKS]       BL50002B Src homology 3 (SH3) domain proteins profile
[SUPFAM]       SH3 homology 6e-17
[PROSITE]      MYRISTYL 6
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 6
[PROSITE]      PKC_PHOSPHO_SITE 8
[PROSITE]      ASN_GLYCOSYLATION 1
[PFAM]         Src homology domain 3
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY 24.04 %

```

```

SEQ  MAELQQLQEF EIPTGREALRGNHSALLRVADYCEDNYVQATDKQKALEETMAFTTQALAS
SEG  .....
laboA .....

SEQ  VAYQVGNLAGHTLRMLDLQGAALRQVEARVSTLGMVMNMHMEKVARREIGTLATVQRLPP
SEG  .....
laboA .....

```



```

SEQ      GQKVIAPENLPPLTPYCRRLNFGCLDDIGHGKDLSTQLSRTGTLRSKSIKAPATPASA
SEG      .....
laboA    .....

SEQ      TLGRPPRIPEPVHLPVVPDGRLSAASSASSLASAGSAEGVGGAPTPKGQAAPPAPPLPSS
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
laboA    .....

SEQ      LDPPPPPAAVEVFQRPPTLEELSPPPPDEELPLPLDLPPPPPLDGDGLPPPPPGFGPD
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
laboA    .....

SEQ      EPSWVPASYLEKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSEGTGFFPGN
SEG      XX.....
laboA    .....EECCCBCCCTTBCCBTTEEEEEEEETTTTEEEEEETTEEEEEEGG

SEQ      YVEPSC
SEG      .....
laboA    GEEE..

```

Prosites for DKFZphfd2_24n20.3

PS00001	22->26	ASN GLYCOSYLATION	PDOC00001
PS00004	339->343	CAMP PHOSPHO_SITE	PDOC00004
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	338->341	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	321->325	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	215->221	MYRISTYL	PDOC00008
PS00008	332->338	MYRISTYL	PDOC00008

Pfam for DKFZphfd2_24n20.3

```

HMM_NAME      Src homology domain 3

HMM            *pyVIALYDYqAqdpDELSFkEGDIIIIIEdsDD.WWrgRnnnTNGQEGW
               ++V+ LY+Y++Q ++ELSF EG +I + + D W++G + +G+
Query          311 EKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSE---GTGF 356

HMM            IPSNYVEPi*
               +P NYVEP
Query          357 FPGNYVEPS 365

```


DKFZphfkd2_24p5

group: intracellular transport and trafficking

DKFZphfkd2_24p5 encodes a novel 811 amino acid protein which is a novel splice variant of human ankyrin G.

The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments.

The new protein can find application in modulating the structure and membrane topology of Ranvier nodes and other neuronal cell membranes.

Human ankyrin G (ANK-3) new splice variant

splice variant

potential frame shift at 2720 was checked
see BLASTX

Sequenced by EMBL

Locus: /map="10q21"

Insert length: 3470 bp

Poly A stretch at pos. 3459, no polyadenylation signal found

```

1  AGCTTTAAAA GGATGTCTGC GAAGTGGTCA AAAGGATCTT AACCTCAATT
51  AAGTGGGGTT TTTTAAAAAG ATTTTGTGGG GGGCCTGAAA TTTTGAAAAT
101 CTTCGAACCTC TGAGTGGGGA AAGATGTATA ATTCCTCAAT TGCCTACGAG
151 GATATCAAGA TGCTGAGAGG AATTCAGCGG TGGTGAAGAG AGTGGATACA
201 AACCAAGGAT TGGTTTCCTT GAGCTGTTTT GGAGGTTGAT TCTAAATCAC
251 TGCTTAAGGA ATTCCTGGAA ACATCAGGAA AACATTTGAT CATCCAAGCC
301 TAGTGGAAAT GGCTTTACCG CAGAGTGAAG ATGCAATGAC CGGGGACACA
351 GACAAATATC TTGGGCCACA GGACCTTAAG GAATTGGGTG ATGATTCCCT
401 GCCTGCAGAG GGTACATGG GCTTTAGTCT CGGAGCGCGT TCTGCCAGCC
451 TCCGCTCCTT CAGTTCGGAT GGGTCTTACA CCTTGAACAG AAGCTCCTAT
501 GCACGGGACA GCATGATGAT TGAAGAACTC CTCGTGCCAT CCAAAGAGCA
551 GCATCTAACA TTCACAAGGG AATTTGATTC AGATTCTCTT AGACATTACA
601 GCTGGGCTGC AGACACCTTA GACAATGTCA ATCTTGTTCC AAGCCCCATT
651 CATCTCTGGT TTCTGGTTAG CTTTATGGTG GACGCGAGAG GGGGCTCCAT
701 GAGAGGAAGC CGTCATCAGC GGATGAGAAT CATCATTCCT CCACGCAAGT
751 GTACGGCCCC CACTCGAATC ACCTGCCGTT TGGTAAAGAG ACATAAACTG
801 GCCAACCACAC CCCCATGGT GGAAGGAGAG GGATTAGCCA GTAGGCTGGT
851 AGAAATGGGT CCTGCAGGGG CACAATTTTT AGGCCCTGTC ATAGTGAAAA
901 TCCCTCACTT TGGGTCCATG AGAGGAAAAG AGAGAGAACT CATTGTCTCT
951 CGAAGTGAAG ATGGTGAAAC TTGGAAGGAG CATCAGTTTG ACAGCAAAAA
1001 TGAAGATTTA ACCGAGTTAC TTAATGGCAT GGATGAAGAA CTTGATAGCC
1051 CAGAAGAGTT AGGGAAAAAG CGTATCTGCA GGATTATCAC GAAAGATTTT
1101 CCCCAGTATT TTGCAGTGGT TTCCCGGATT AAGCAGGAAA GCAACCAGAT
1151 TGGTCCCTGAA GGTGGAATTC TGAGCAGCAC CACAGTGCCC CTTGTTCAG
1201 CATCTTTCCC AGAGGGTGCC CTAACATAAA GAATTCGAGT GGGCCTCCAG
1251 GCCCAGCCTG TTCCAGATGA AATTGTGAAA AAGATCCTTG GAAACAAAGC
1301 AACTTTTAGC CCAATTGTCA CTGTGGAACC AAGAAGACGG AAATTCATA
1351 AACCAATCAC AATGACCATT CCGGTGCCCC CGCCCTCAGG AGAAGGTGTA
1401 TCCAATGGAT ACAAAGGGGA CACTACACCC AATCTGCGTC TTCTCTGTAG
1451 CATTACAGGG GGCACCTTCG CTGCTCAGTG GGAAGACATC ACAGGAACAA
1501 CTCCTTTGAC GTTTATAAAA GATTGTGTCT CCTTTACAAC CAATGTTTCA
1551 GCCAGATTTT GGCTTGCAGA CTGCCATCAA GTTTTAGAAA CTGTGGGGTT
1601 AGCCACGCAA CTGTACAGAG AATTGATATG TGTTCCATAT ATGGCCAAGT
1651 TTGTTGTTTT TGCCAAAATG AATGATCCCG TAGAATCTTC CTTGCCATGT
1701 TTCTGCATGA CAGATGACAA AGTGGACAAA ACTTTAGAGC AACAAGAGAA
1751 TTTTGAGGAA GTCGCAAGAA GCAAAGATAT TGAGGTTCTG GAAGGAAAAC
1801 CTATTTATGT TGATTGTTAT GGAATTTGG CCCCCTTAC CAAAGGAGGA
1851 CAGCAACTGT TTTTAACTT TTATTCTTTC AAAGAAAAAT GACTGCCATT
1901 TTCCATCAAG ATTAGAGACA CCAGCCAAGA GCCCTGTGGT CGTCTGTCTT
1951 TTCTGAAAGA ACCAAAGACA ACAAAGGAC TGCCTCAAAC AGCGGTTTGC
2001 AACTTAAATA TCACTCTGCC AGCACATAAA AAGATTGAGA AAACAGATGG
2051 ACGACAGAGC TTCGATCCT TAGCTTTACG TAAGCGCTAC AGCTACTTGA
2101 CTGAGCCTGG AATGAGTCCA CAGAGTCCAT GTGAACGGAC AGATATCAGG
2151 ATGGCAATAG TAGCCGATCA CCTGGGACTT AGTTGGACAG AACTGGCAAG
2201 GGAAGCTGAAT TTTTCAGTGG ATGAAATCAA TCAAATACGT GTGGAAAATC
2251 CAAATCTTTT AATTCTCAG AGCTTCATGT TTTTAAAAAA ATGGGTTACC
2301 AGAGACGGAA AAAATGCCAC AACTGATGCC TTAACCTCGG TCTTGACAAA
2351 AATTAATCGA ATAGATATAG TGACACTGCT AGAAGGACCA ATATTGATT

```



```

2401 ATGGAAATAT TTCAGGCACC AGAAGTTTTC CAGATGAGAA CAATGTTTTTC
2451 CATGACCCCTG TTGATGGTTA TCCTTCCCTT CAAGTGGAAAC TGGAAACCCC
2501 CACAGGGTTG CACTACACAC CACCTACCCC TTTCCAGCAA GATGATTATT
2551 TTAGTGATAT CTCTAGCATA GAATCTCCCC TTAGAACCCC TAGTAGACTG
2601 AGTGATGGGC TAGTGCCTTC CCAGGGGAAC ATAGAGCATT CCGCAGATGG
2651 ACCTCCAGTC GTAACCTGCAG AAGACGCTTC CTTAGAAGAC AGCAAACTGG
2701 AAGACTCACT GCCTTTAACA GAAATGCCTG AAGCAGTGAT GTAGATGAGA
2751 GCCAGTTGGA GAATGTATGT CTGAGTTGGC AGAATGAGAC ATCAAGTGGA
2801 AACCTAGAGT CTTGCGCTCA AGCTCGAAGA GTAACCTGGT GGTACTAGA
2851 TCGACTGGAT GACAGCCCTG ACCAGTGTAG AGATTCCATT ACCTCATATC
2901 TCAAAGGAGA AGCTGGCAAA TTTGAAGCAA ATGGAAGCCA TACAGAAATC
2951 ACTCCAGAAG CAAAGACAAA ATCTTACTTT CCAGAAATCCC AAAATGATGT
3001 AGGAAAACAG AGTACCAAGG AACTCTGAA ACCAAAAATA CATGGATCTG
3051 GTCATGTTGA AGAACCAGCA TCACCACTAG CAGCATATCA GAAATCTCTA
3101 GAAGAAACCA GCAAGCTTAT AATAGAAGAG ACTAAACCCCT GTGTGCCTGT
3151 CAGTATGAAA AAGATGAGTA GGAATCTCC AGCAGATGGC AAGCCAAGGC
3201 TTAGCCTCCA TGAAGAAGAG GGGTCCAGTG GGTCTGAGCA AAGCAGGGA
3251 GAAGGTTTTA AGGTGAAAAAC GAAGAAAGAA ATCCGGCATG TGGAAAAGAA
3301 GAGCCACTCG TAACAGCGAA CGGTCAGTCA AGGATCATAA GTTTTACTG
3351 CCAGTATTGA GAAATTCGTG GAAGAAATGT CAGCAGGAAG TAAAAATTCA
3401 CCGAGAAGTG TGTGTGTGTT CGTGCTTCC ACACATTAAT GGCATGATT
3451 TTTTATGCA AAAAAAAAAA

```

BLAST Results

Entry MMANK3A_1 from database TREMBL:
Ank3"; product: "ankyrin 3"; Mus mu... +3 4022 0.0 2

Entry HS13616 from database EMBL:
Human ankyrin G (ANK-3) mRNA, complete cds.
Length = 14,770
Plus Strand HSPs:
Score = 8505 (1276.1 bits), Expect = 0.0, Sum P(3) = 0.0
Identities = 1799/1873 (96%)

Medline entries

95394457:
Chromosomal localization of the ankyrinG gene
(ANK3/Ank3) to human 10q21 and mouse 10.

95138209:
A new ankyrin gene with neural-specific isoforms localized at the
axonal initial segment and node of Ranvier

Peptide information for frame 3

ORF from 309 bp to 2741 bp; peptide length: 811
Category: known protein
Classification: unset

```

1 MALPQSEDAM TGDIDKYLGP QDLKELGDDS LPAEGYMGFS LGARSASLRS
51 FSSDGSYTLN RSSYARDSMM IEELLVPSKE QHLETFREPD SDSLRHYSWA
101 ADTLDNVNLV PSPIHSGFLV SFMVDARGGS MRGSRHHGMR IIPPRKCTA
151 PTRITCRLVK RHKLNPPEM VEGGLASRL VEMGPAGAOF LGPVIVEIPH
201 FGSMRGKERE LIVLRSENGE TWKEHQFDSK NEDLTELLNG MDEELDSPEE
251 LGKKRICRII TKDFPQYFAV VSRIKQESNQ IGPEGGILSS TTVPLVQASF
301 PEGALTKRIR VGLQAQVPVD EIVKKILGNK ATFSPIVTVE PRRRRFHKPI
351 TMTIPVPPPS GEGVSNYKYG DTTPLNRLLC SITGGTSPAQ WEDITGTTPL
401 TFIKDCVSFT TNVSAREFLA DCHQVLETVG LATQLYRELI CVPYMAKFVV
451 FAKMNDPVES SLRCFCMTDD KVDKTLEQQE NFEEVARSKD IEVLEKPIY
501 VDCYGNLAPL TKGGQQLVFN FYSFKENRLP FSIKIRDTSQ EPCGRLSFLK
551 EPKTTKGLPQ TAVCNLNLITL PAHKKIEKTD GRQSFASLAL RKRYSYLTEP
601 GMSQSPSCER TDIRMAIVAD HLGLSWTELA RELNFSVDEI NQIRVENPNPS
651 LISQSFMLK KWTTRDGKNA TTDALTSVLT KINRIDIVTL LEGPIFDYGN
701 ISGTRSFAD NNVFHDVPDG YPSLQVELET PTGLHYTPPT PFQDDYFSD
751 ISSIESPLRT PSRLSDGLVP SQGNIEHSAD GPPVVTAEDA SLEDSKLEDS
801 VPLTEMPEAV M

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phkd2_24p5, frame 3

TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds., N = 1, Score = 4022, P = 0

TREMBL:MMANK3B_3 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

TREMBL:MMANK3B_4 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

>TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds.
Length = 1,094

HSPs:

Score = 4022 (603.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 769/805 (95%), Positives = 783/805 (97%)

```

Query:      1 MALPQSEDAMTGDGTDKYLGPQDLKELGDDSLPAEGYMGFSLGARSASLRSFSSOGSYTLN 60
MALP SEDA+TGDTDKYLGPDQLKELGDDSLPAEGY+GFSLGARSASLRSFSSD SYTLN
Sbjct:      1 MALPHSEDAITGDTDKYLGPQDLKELGDDSLPAEGYVGFSLGARSASLRSFSSDRSYTLN 60

Query:     61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV 120
RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLV SP+HSGFLV
Sbjct:     61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVSSPVHSGFLV 120

Query:    121 SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180
SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
Sbjct:    121 SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180

Query:    181 VEMGPAGAQLGPPVIVEIPHEGSMRGKERELIVLRSENGETWKEHQFDSKNEDELTELLNG 240
VEMGPAGAQLGPPVIVEIPHEGSMRGKERELIVLRSENGETWKEHQFDSKNEDEL ELLNG
Sbjct:    181 VEMGPAGAQLGPPVIVEIPHEGSMRGKERELIVLRSENGETWKEHQFDSKNEDELLAELLNG 240

Query:    241 MDEELDSPEELGKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300
MDEELDSPEELG KRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
Sbjct:    241 MDEELDSPEELGKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300

Query:    301 PEGALTKRIRVGLQAQVPVDEIVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS 360
PEGALTKRIRVGLQAQVPV+E VKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS
Sbjct:    301 PEGALTKRIRVGLQAQVPVETVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS 360

Query:    361 GEGVSNYKGDTPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420
GEGVSNYKGD TPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
Sbjct:    361 GEGVSNYKGDATPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420

Query:    421 DCHQVLETVGLATQLYRELICVPYMAKFVVFVAKMNDPVESLRCFCMTDDKVDKTEQQE 480
DCHQVLETVGLA+QLYRELICVPYMAKFVVFVAK NDPVESLRCFCMTDD+VDKTEQQE
Sbjct:    421 DCHQVLETVGLASQLYRELICVPYMAKFVVFVAKTNDPVESLRCFCMTDDRVDKTEQQE 480

Query:    481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSEKFNRLPFSIKIRDTSQ 540
NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSEKFNRLPFSIKIRDTSQ
Sbjct:    481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSEKFNRLPFSIKIRDTSQ 540

Query:    541 EPCGRSLFLEKPKTTKGLPQTAVCNLNLITLPAHKKIEKTQGRQSFASLALRKRYSYLTEP 600
EPCGRSLFLEKPKTTKGLPQTAVCNLNLITLPAHKK EK D RQSFASLALRKRYSYLTEP
Sbjct:    541 EPCGRSLFLEKPKTTKGLPQTAVCNLNLITLPAHKKAEKADRRQSFASLALRKRYSYLTEP 600

Query:    601 GMSQPSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLK 660
GMSQPSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSF LK
Sbjct:    601 SMSQPSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLLK 660

Query:    661 KWTTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG 720
KWTTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG
Sbjct:    661 KWTTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG 720

Query:    721 YPSLQVELETPTGLHYTPPTPFQQDDYFSDISSIESPLRTPSRSLDGLVPSQGNIEHSAD 780
+PS QVELETG GL++TP PFQQDD+FSDISSIESP RTPSRSLDGLVPSQGNIEH
Sbjct:    721 HPSFQVELETPMGLYWTTPNPFQQDDHFSDISSIESPFTPSRSLDGLVPSQGNIEHPTG 780

Query:    781 GPPVVTAEDASLEDSKLEDSVPLTE 805
GPPVVTAED SLEDSK++DSV +T+

```


Sbjct: 781 GPPVVTAEDTSLEDSKMDDSVTVTD 805

Pedant information for DKFZphfkd2_24p5, frame 3

Report for DKFZphfkd2_24p5.3

[LENGTH] 811
[MW] 90104.66
[pI] 5.40
[HOMOL] TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial
ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds. 0.0
[BLOCKS] BL500178 Death domain proteins profile
[PIRKW] phosphoprotein: 0.0
[PIRKW] alternative splicing 0.0
[PIRKW] peripheral membrane protein 0.0
[PIRKW] cytoskeleton 0.0
[SUPFAM] ankyrin 0.0
[SUPFAM] ankyrin repeat homology 0.0
[SUPFAM] unassigned ankyrin repeat proteins 0.0
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 1.73 %

SEQ MALPQSEDAMTGDTKYLGPDQLKELGDDSLPAEGYMGFSLGARSASLSRFSSDGSYTLN
SEG
PRD ccc
MEM

SEQ RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV
SEG
PRD cccchhhhhhhhhheeeehhhhhhhhhhhcccccccccccccccccccccccccccccccccc
MEMMMMMMMMMMM

SEQ SFMVDARGGSMRGRHGMRIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
SEG
PRD eeeeecc
MEM MMMMMMMMMMMMMMM.....M

SEQ VEMGPAGAQLGPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG
SEG
PRD ecc
MEM MMM

SEQ MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
SEG
PRD cccccchhhhhhhhhheeecc
MEM

SEQ PEGALTKRIRVGLQAQVPDEIVKKILGNKATFSPIVTVEPRRRKFHKBITMTIPVPPPS
SEG
PRD ccchhhhhhhhhhhhhcc
MEM

SEQ GEGVSNYKGDTPNRLRLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
SEG
PRD ccc
MEM

SEQ DCHQVLETVGLATQLYRELICVPYMAKFVVFAMNDPVESSLRCFCMTDDKVDKTLQQE
SEG
PRD cchhh
MEM

SEQ NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ
SEG
PRD ccc
MEM

SEQ EPCGRSLFLKEPKTKGLPQTAVCNLNLITLPAHKKIEKTPGRQSFASLALRKRYSYLTP
SEG
PRD ccc
MEM

SEQ GMSPOSPCERDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMLK
SEG
PRD cccccchhh
MEM


```

SEQ      KQVTRDQGNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVDHDPVQD
SEG      .....
PRD      hhhhhccccccchhhhhhhhhccceeeeecccccccccccccccccccccccccccccc
MEM      .....

SEQ      YPSLQVELETPTGLHYTPPTPFQDDYFSDISSIESPLRTPSRSLDGLVPSQGNIEHSAD
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      GPPVVTAEASLEDSKLEDSVPLTEMPEAVM
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccc
MEM      .....

```

(No Prosite data available for DKFZphfkd2_24p5.3)

(No Pfam data available for DKFZphfkd2_24p5.3)

DKF2phfkd2_3i13

group: transmembrane protein

DKF2phfkd2_3i13 encodes a novel 406 amino acid protein with *C. elegans* cosmid Y37D8A and *A. thaliana* H71412 hypothetical protein.

The novel protein contains 3 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to *A.thaliana* and *C.elegans*;
membrane regions: 3

complete cDNA, complete cds, EST hits

Sequenced by BMF2

Locus: /map="17"

Insert length: 2052 bp
Poly A stretch at pos. 2032, no polyadenylation signal found

```
1 AGTGACGTGA GCGGGTTCCG GTTGTCTGGA GCCCAGCGGC GGGTGTGAGA
51 GTCCGTAAGG AGCAGCTTCC AGGATCCTGA GATCCGGAGC AGCCGGGGTC
101 GGAGCGGCTC CTCAAGAGTT ACTGATCTAT GAAATGGCAG AGAATGGAAA
151 AAATTGTGAC CAGAGACGTG TAGCAATGAA CAAGGAACAT CATAATGGAA
201 ATTTACACAGA CCCCTCTTCA GTGAATGAAA AGAAGAGGAG GGAGCGGGAA
251 GAAAGGCAGA ATATTGTCTT GTGGAGACAG CCGCTCATT CCTTGACGTA
301 TTTTCTCTCT GAAATCCTTG TAATCTTGAA GGAATGGACC TCAAAATTAT
351 GGCATCGTCA AAGCATTGTG GTGTCTTTT TACTGCTGCT TGCTGTGCTT
401 ATAGCTACGT ATTATGTTGA AGGAGTGCAT CAACAGTATG TGCAACGTAT
451 AGAGAAACAG TTTCTTTTGT ATGCCTACTG GATAGGCTTA GGAATTTTGT
501 CTTCTGTGGG GCTTGGAAAC GGGCTGCACA CTTTCTGCT TTATCTGGGT
551 CCACATATAG CCTCAGTTAC ATTAGCTGCT TATGAATGCA ATTCAGTTAA
601 TTTTCCCGAA CCACCCATAT CTGATCAGAT TATTGTCCA GATGAAGAGG
651 GCACTGAAGG AACCATTTTT TTGTGGAGTA TCATCTCAAA AGTTAGGATT
701 GAAGCCTGCA TGTGGGGTAT CGGTACAGCA ATCGGAGAGC TGCCTCCATA
751 TTTTCATGCC AGAGCAGCTC GCCTCTCAGG TGCTGAACCA GATGATGAAG
801 AGTATCAGGA ATTTGAAGAG ATGCTGGAAC ATGCAGAGTC TGCACAAGAC
851 TTTTGCCCTCC GGGCCAAACT GGCAGTTCAA AAACAGTAC AGAAAGTTGG
901 ATTTTGTGGA ATTTGGCCT GTGCTTCAAT TCCAAATCCT TTATTGTATC
951 TGGCTGGAAT AACGTGTGGA CACTTCTGG TACCTTTTGT GACCTTCTTT
1001 GGTGCAACCC TAATTGGAAA AGCAATAATA AAAATGCATA TCCAGAAAAT
1051 TTTTGTGTTA ATAACATTCA GCAAGCACAT AGTGGAGCAA ATGGTGGCTT
1101 TCATTGGTGC TGTCCCCGGC ATAGGTCCAT CTCTGCAGAA GCCATTTCAG
1151 GAGTACCTGG AGGCTCAACG GCAGAAGCTT CACCACAAA GCGAAATGGG
1201 CACACCCACG GGAGAAAACG GGTGTCTCTG GATGTTTGAA AAGTTGGTCG
1251 TTGTCATGGT GTGTTACTTC ATCCTATCTA TCATTAATCT CATGGCACA
1301 AGTTATGCCA AACGAATCCA GCAGCGGTTG AACTCAGAGG AGAAAATAA
1351 ATAAGTAGAG AAAGTTTTAA ACTGCAGAAA TTGGAGTGGG TGGGTCTGTC
1401 CTTAAATTGG GAGGACTCCA AGCCGGGAGG GAAAATTCCT TTTTCCAACC
1451 TGTATCAATT TTTACAACCT TTTTCTGAA AGCAGTTTAG TCCATACTTT
1501 GCACTGACAT ACTTTTCTCT TCTGTGCTAA GGTAAGGTAT CCACCTCGA
1551 TGCAATCCAC CTTGTGTTTT CTAGGGTGG AATGTGATGT TCAGCAGCAA
1601 ACTTGCAACA GACTGGCCTT CTGTTTGTTA CTTTCAAAG GCCCAGATGA
1651 TACAATTAGA GAATCCACAC CGCACAAAAA AAGTTCCTAA GTATGTTAAA
1701 TATGTCAAGC TTTTATAGCT TGTCACAAAT GATTGCTTTG TTTTCTTAAG
1751 TCATCAAAAT GTATATAAAT TATCTAGATT GGATAACAGT CTTGCATGTT
1801 TATCATGTTA CAATTTAATA TTCCATCCTG CCAACCCCTT CCTCTCCCAT
1851 CCTCAAAAAA GGGCCATTTT ATGATGCATT GCACACCCCT TGGGGAAATT
1901 GATCTTTAAA TTTTGAGACA GTATAAGGAA AATCTGGTTG GTGCTTTACA
1951 AGTGAGCTGA CACCATTTTT TATTCTGTGT ATTTAGGATG AAGTCTTGAA
2001 AAAAACTTTA TAAAGACATC TTTAATCATT CCAAAAAAAA AAAAAAAA
2051 AA
```

BLAST Results

Entry AC004686 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone
hRPC.1073_F15; HTGS phase 1, 8 unordered pieces.
Score = 4142, P = 6.1e-199, identities = 830/832

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 1351 bp; peptide length: 406
 Category: similarity to unknown protein

```

1 MAENGKNCDO RRVAMNKEHH NGNFTDPSSV NEKKRREREE RQNIWLWRQP
51 LITLQYFSLE ILVILKEWTS KLWHRQSIVV SFLLLAVLI ATYYVEGVHQ
101 QYVQRIEQF LLYAYWIGLG ILSSVGLGTG LHTFLLYLGP HIASVTLAAY
151 ECNSVNFPEP PYPDQIICPD EEGTEGTIFL WSIISKVRIE ACMWGIGTAI
201 GELPPYFMMAR AARLSGAEPD DEEYQEFEEF LEHAESAQDF ASRAKLAVQK
251 LVQKVGFVFI LACASIPNPL FDLGITCGH FLVPFWTFGG ATLIGKAIK
301 MHIQKIFVII TFSKHIVEQM VAFIGAVPGI GPSLQKPFQE YLEAQRQKLH
351 HKSEMGTPOG ENWLSWMFEK LVVVMVCYFI LSIINSMAQS YAKRIQORLN
401 SEEKTK
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_3i13, frame 2

TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid
 Y37D8A, N = 1, Score = 905, P = 8.8e-91

TREMBL:ATAC98_2 gene: "YUP8H12.2"; Arabidopsis thaliana chromosome 1
 YAC yUP8H12 complete sequence., N = 1, Score = 470, P = 1.1e-44

PIR:H71412 hypothetical protein - Arabidopsis thaliana, N = 1, Score =
 293, P = 6e-24

>TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid
 Y37D8A
 Length = 457

HSPs:

Score = 905 (135.8 bits), Expect = 8.8e-91, P = 8.8e-91
 Identities = 167/317 (52%), Positives = 228/317 (71%)

```

Query:   38 REERQNIWLWRQPLITLQYFSLEILVILKEWTSKLWHRQSIVVSFLLLAVLIATYYVEG 97
          R ER+ IV WR+P I + Y +EI + E K+ +++++ + + + + Y+ G
Sbjct:   93 RMERETIVFWRRPHIVIPYALMEIAHLAVELFFKILAHKTVLLLTASISGLAVYGYHAPG 152

Query:   98 VHQQYVQRIEQFLLYAYWIGLGILSSVGLGTGLHTFLLYLGP HIASVTLAAYECNSVNF 157
          HQ++VQ IEK L +++W+ LG+LSS+GLG+GLHTFL+YLGPHIA+VT+AAVEG S++F
Sbjct:   153 AHQEHVQTIEKHILWWSWVLLGVLLSSIGLGSGLHTFLIYLGPHIAAVTMAAYECQSLDF 212

Query:   158 PEPYPYDQIICPDDEEGTEGTIFLWSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGA 217
          P+PPYP+ I CP + + F W I++KVR+E+ +WG GTA+GELPPYFMARAAR+SG
Sbjct:   213 PQPPYPESIQCFSTKSSIAVTF-WQIVAKVRVESLLWGAGTALGELPPYFMARAARISGQ 271

Query:   218 EPDDEEYQEFEEMLE-HAESAQD----FASRAKLAVQKLVQKVGFFGILACASIPNPLFD 272
          EPDDEEY+EF E++ ES D RAK V+ + ++GF GIL ASIPNPLFD
Sbjct:   272 EPDDEEYREFLELMNADKESDADQKLSIVERAKSWVEHNIHRLGFP GILLFASIPNPLFD 331

Query:   273 LAGITCGHFLVPFWTFFGATLIGKAIKMHQKIFVIITFSKHIVEQMVAFIGAVPGIGP 332
          LAGITCGHFLVPFW+FFGATLIGKA++KMH+Q FVI+ FS H E V + +P +GP
Sbjct:   332 LAGITCGHFLVPFWTFFGATLIGKALVKMHVQMGEVILAFSDHHAENFVKILEKIPAVGP 391

Query:   333 SLQKPFQEYLEAQRQKLH 350
          +++P + LE QR+ LH
Sbjct:   392 YIRQPISDLLEKQRKALH 409
  
```

Pedant information for DKFZphfkd2_3i13, frame 2

Report for DKFZphfkd2_3i13.2


```

SEQ      MAENGNKCDQRRVAMNKEHHNGNFTDPSSVNEKKRREREERQNIVLWRQPLITLQYFSLX
SEG      .....xxxxxxxxxxxxx.....
PRD      cccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhccccchhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      ILVILKEWTSKLWHRQSIIVSVFLLLLAVLAIATYVVEGVHQQYVQRIEKQFLLYAYWIGLG
SEG      .....xxxxxxxxxxxxx.....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhhhhhhhh
MEM      MM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      ILSSVGLGTGLHTFLLYLGPFIASVTLAAYECNSVNFPEPPYPDQIICPDEEGTGTIFL
SEG      xxxxxxxxxxxxxxxx.....
PRD      hccccccccceeeeeeeccchhhhhhhhhhhhhccccccccccccccccccccccccceeee
MEM      .....

SEQ      WSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGAEPDDEEYQEFEEMLEHAESAQDF
SEG      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      eehhhhhhhhhhhhhhhccccccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      ASRAKLAVQKLQKVGFFGILACASIPNPLFDLAGITCGHFLVPFWTFFGATLIGKAIK
SEG      .....
PRD      hhhhhhhhhhhhhhhhhccccccccccccccccccccccccceeeeeeeehhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      MHIQKIFVITITFSKHIVEQMVAFIGAVPGIGPSLQKPFQEYLEAQROKLHHKSEMGTPOG
SEG      .....
PRD      hhhhheeeeeeechhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhcccccccc
MEM      .....

SEQ      ENWLSWMFEKLVVVMVCYFILSIINSMAQSYAKRIQRLNSEETK
SEG      .....
PRD      cchhhhhhhhhhhheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc
MEM      .....

```

PS000001	23->27	ASN_GLYCOSYLATION	PDOC000001
PS000005	69->72	PKC_PHOSPHO_SITE	PDOC000005
PS000006	29->33	CK2_PHOSPHO_SITE	PDOC000006
PS000006	215->219	CK2_PHOSPHO_SITE	PDOC000006
PS000006	236->240	CK2_PHOSPHO_SITE	PDOC000006
PS000008	120->126	MYRISTYL	PDOC000008
PS000008	126->132	MYRISTYL	PDOC000008
PS000008	173->179	MYRISTYL	PDOC000008
PS000008	195->201	MYRISTYL	PDOC000008
PS000008	197->203	MYRISTYL	PDOC000008
PS000008	259->265	MYRISTYL	PDOC000008
PS000008	275->281	MYRISTYL	PDOC000008
PS000008	325->331	MYRISTYL	PDOC000008
PS000008	329->335	MYRISTYL	PDOC000008
PS000008	356->362	MYRISTYL	PDOC000008

399

DKFZphfkd2_3ol7

group: metabolism

DKFZphfkd2_3ol7 encodes a novel 72 amino acid protein with similarity to bos taurus NADH-ubiquinone oxidoreductase B33 subunit (EC 1.6.5.3) (EC 1.6.99.3).

NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. The novel protein is the human orthologue of bovine B22.

The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

strong similarity to bovine NADH-UBIQUINONE OXIDOREDUCTASE B22 subunit

complete cDNA, complete cds, EST hits,
in frame stop codon at -274 will be checked
ESTs HS1291620/AA883920 show no stop codon at this side

Sequenced by BMFZ

Locus: unknown

Insert length: 693 bp

Poly A stretch at pos. 670, polyadenylation signal at pos. 659

```

1 CAGCAGGCGT GCAGTTTCCC GGCTCTCCGC GCGGCCGGGG AAGGTCAGCG
51 CCGTAATGGC GTTCTTGCGG TCGGGACCCT ACCTGACCCA TCAGCAAAAG
101 GTGTTGCGGC TTTATAAGCG GCGCTACGC CACCTCGAGT CGTGGTGCCT
151 CCAGAGAGAC AAATACCGAT ACTTTGCTTG TTTGATGAGA GCCCGGTTTG
201 AAGAACATAA GAATGAAAAG GATATGGCGA AGGCCACCCA GCTGCTGAAG
251 GAGGCCGAGG AAGAATTCTG GTAACGTCAG CATCCACAGC CATACATCTT
301 CCCTGACTCT CCTGGGGGCA CCTCCTATGA GAGATACGAT TGCTACAAGG
351 TCCCAGAATG GTGCTTAGAT GACTGGCATC CTTCTGAGAA GGCAATGTAT
401 CCTGATTACT TTGCCAAGAG AGAACAGTGG AAGAACTGC GGAGGGAAG
451 CTGGGAACGA GAGGTTAAGC AGCTGCAGGA GAAACGCCA CCTGGTGGTC
501 CTTTAACTGA AGCTTTGCCC CCGCCCGAA AGGAAGGTGA TTTGCCCCCA
551 CTGTGGTGGT ATATTGTGAC CAGACCCCGG GAGCGGCCCA TGAGAAAGA
601 GAGAGACCTC ATCTTTCATG CTTGCAAGTG AAATATGTTA CAGAACATGC
651 ACTTGCCCTA ATAAAAATC AGTAAAAAAA AAAAAAAAAA AAA

```

BLAST Results

Entry S28256 from database PIR:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
>TREMBL:MIBTCIB22_1 gene: "CI-B22"; product: "NADH-ubiquinone
oxidoreductase complex B22 subunit"; B.taurus mitochondrion CI-B22
mRNA for B22 subunit of the NADH-ubiquinone oxidoreductase complex
Score = 933, P = 5.2e-93, identities = 163/179, positives = 172/179,
frame +2

Medline entries

92389317
Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from RT bovine heart mitochondria.
Application of a novel strategy for RT sequencing proteins using the polymerase chain reaction

Peptide information for frame 2

ORF from 56 bp to 271 bp; peptide length: 72
Category: strong similarity to known protein

```

1 MAFLASGPYL THQQKVLRLY KRALRHLESW CVQDKYRYF ACLMRARFEE
51 HKNEKDMAKA TQLLKEAEEE FW*ROHPQPY IFPDSPPGTS YERYDCYKVP
101 EWCLDDWHPs EKAMYPDYFA KREQWKKLRR ESWEREVKQL QEETPPGGPL
151 TEALPPARKE GDLPLWYI VTRPRERPM

```


BLASTP hits

Sequences producing significant alignments: (bits) Value

sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE... 141 7e-34
 tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO NADH-UBIQ... 53 3e-07

>sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE
 OXIDOREDUCTASE B22 SUBUNIT (EC 1.6.5.3) (EC 1.6.99.3)
 (COMPLEX I-B22) (CI-B22).[BOS TAURUS]
 Length = 178

Score = 141 bits (351), Expect = 7e-34
 Identities = 63/71 (88%), Positives = 68/71 (95%)

Query: 2 AFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKAT 61
 AFL+SG YLTHQQKVLRLYKRALRHLESWC+ RDKYRYFACL+RARF+EHKNEKDM KAT
 Sbjct: 1 AFLSSGAYLTHQQKVLRLYKRALRHLESWCIHRRDKYRYFACLLRARFDEHKNEKDMVKAT 60

Query: 62 QLLKEAEEEFW 72
 QLL+EAEEEFW
 Sbjct: 61 QLLREAEEEFW 71

>tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO
 NADH-UBIQUINONE OXIDOREDUCTASE B22.[CAENORHABDITIS
 ELEGANS]
 Length = 163

Score = 52.7 bits (124), Expect = 3e-07
 Identities = 25/64 (39%), Positives = 41/64 (64%), Gaps = 1/64 (1%)

Query: 10 LTHQQKVLRLYKRALRHLESWCVQRD-KYRYFACLMRARFEEHKNEKDMAKATQLLKEAE 68
 L+H+QKV RLYKR LR +++W + + R+ C++RARF+ + +E D K+ LL +
 Sbjct: 12 LSHRQKVTRLYKRCLEVDNWNWYGGNNLEVRQKCIIRARFDANAEVDTRKSQILLADGC 71

Query: 69 EEFW 72
 + W
 Sbjct: 72 RQLW 75

Alert BLASTP hits for DKFZphfkd2_3o17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_3o17, frame 2

Report for DKFZphfkd2_3o17.2

[LENGTH] 72
 [MW] 8839.28
 [pI] 9.26
 [HOMOL] PIR:S28256 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
 2e-34
 [KW] All_Alpha

SEQ MAFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKA
 PRD ccc

SEQ TQLLKEAEEEFW
 PRD hhhhhhhhhccc

(No Prosite data available for DKFZphfkd2_3o17.2)

(No Pfam data available for DKFZphfkd2_3o17.2)

DKF2phfkd2_46a6

group: kidney derived

DKF2phfkd2_46a6 encodes a novel 315 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="228.6 cR from top of Chr15 linkage group"

Insert length: 2774 bp

Poly A stretch at pos. 2751, polyadenylation signal at pos. 2732

```
1 CTCGCGAGCG CAGCTATGGC TGCTGGCGTA CCCTGTGCGT TAGTCACCAG
51 CTGCTCCTCC GTCTTCTCAG GAGACCAGCT GGTCCAACAT ACCCTTGGAA
101 CAGAAGATCT TATTGTGGAA GTGACTTCCA ATGATGCTGT GAGATTTTAT
151 CCCTGGACCA TTGATAATAA ATACTATTCA GCAGACATCA ATCTATGTGT
201 GGTGCCAAAC AAATTTCTTG TTAGTGCAGA GATTGCAGAA TCTGTCCAAG
251 CATTTGTGGT TTAGTTTGAC AGCACACGAA AATCGGGCCT TGATAGTGTG
301 TCCTCATGGC TTCCACTGGC AAAAGCATGG TTACCTGAGG TGATGATCTT
351 GGTCTGCGAT AGAGTGTCTG AAGATGGTAT AAACCGACAA AAAGCTCAAG
401 AATGGAGCCT CAAACATGGC TTTGAATTGG TAGAACTTAG TCCAGAGGAG
451 TTGCCTGAGG AGGATGATGA CTTCCCAGAA TCTACAGGAG TAAAGCGAAT
501 TGTCCAAGCC CTGAATGCCA ATGTGTGGTC CAATGTAGTG ATGAAGAATG
551 ATAGGAACCA AGGCTTTAGC CTTCTCAACT CATTGACTGG AACAAACCAT
601 AGCATTGGGT CAGCAGATCC CTGTCAACCA GAGCAACCCC ATTTGCCAGC
651 AGCAGATAGT ACTGAATCCC TCTCTGATCA TCGGGGTGGT GCATCTAACA
701 CAACAGATGC CAGGTTGAT AGCATTGTGG ATCCCATGTT AGATCTGGAT
751 ATTCAAGAAT TAGCCAGTCT TACCACTGGA GGAGGAGATG TGGAGAATTT
801 TGAAGACCCC TTTTCAAAGT TAAAGGAAAT GAAAGACAAG GCTGCGACGC
851 TTCTCATGTA GCAAGAAAAA GTGCATGCAG AAAAGGTGGC CAAAGCATTC
901 TGGATGGCAA TCGGGGGAGA CAGAGATGAA ATTGAAGGCC TTTTCTCTGA
951 TGGAGAGCAC TGAATTATTC ATACTAGGGT TTGACCAACA AAGATGCTAG
1001 CTGTCTCTGA GATACCTCTC TACTCAGCCC AGTCATATTT TGCCAAAATT
1051 GCCCTTATCA TGTGGCTGCG CTGACTTGTT TATAGGGTCC CCTTAATTTT
1101 AGTTTTTAGT AGGAGGTAA GGAGAAATCT TTTTTTCTCT CAGTATATTG
1151 TAAGAGAGTG AGGAATACAG TGATAGTAAT GAGTGAGGAT TTCTTAAATA
1201 TACTTTTTTT TTGTTCTAGG AATGAGGGTA GGATAAATCT CAGAGGTCTG
1251 TGTGATTATG TCAAGTTGAA GACAACCTCC AGGCCATTCC TGGTCAACCT
1301 TTTAAGTAGC ATTTCCAGCA TTCACACTTG ATACTGCACA TCAGGAGTTG
1351 TGTCACCTTT CCTGGGTGAT TTGGGTTTTC TCCATTCAAG GAGCTGTAG
1401 CTCTGAGCTA TGATGCTTTT ATTGGGAGGA AAGGAGGCAG CTGCAGAATT
1451 GATGTGAGCT ATGTGGGGCC GAAGTCTCAG CCCGCAGCTA AGTCTCTACC
1501 TAAGAAAATG CCTCTGGGCA TTCTTTTGAA GTATAGTGTG TGAGCTCATG
1551 CTAGAAAAGAA TCAAAAAGCC AGTGTGGATT TTTAGGCTGT AATAAATGAG
1601 GCAAAGGATT TCTATTCCAG TGGGAAGGAA ACCTCTCTAC TGAGTTGTGG
1651 GGGATATGTT GTATGTTAGA GAGAACCTTA AGGAGTCCTT GTATGGGCCA
1701 TGGAGACAGT ATGTGATAAC ATACCGTGAT TTTTATGAAG AAATTCTTCT
1751 GTCTTAGAGT TCTCCCCTGC TGCTTGAGAT GCCAGAGCTG TGTGTTGCA
1801 CACCTGCAAA ACAAGGCACA TTTCCCCCTT TCTCTTAAAG GCCAAAGAGA
1851 GATCACTGCC AAAGTGGGAG CACTAAGGGG TGGGTGGGGA AGTGAAATGT
1901 TAGGCGATGA ATTCTGAGC ACCTTGTTTT TCTTCCAAGG TTCGTAGCTC
1951 CTCTCTGCCC TTCCAAGCCT GTAACCTCGG AGGACTATCT TTTGTTCTCT
2001 ATCCTTTTGT TTGTTAGAGT GGGTCAGCCC CAGAGGAACT GATAAGCAAA
2051 TGGCAAGTTT TTAAGGAAG AGTGGAAAGT ACTGCAAATA AAAATCCTTA
2101 TTTGTTTTTG TAGACTTTGT AATGCATATC ATTAGCCCTC ACTGTGATCA
2151 TTACTGCTGT GGCTCTGAAC TGGCAGATAG TACAGTGGAT GGAAGGTGCC
2201 CGCACACCAG CTGAGAACTG GTTCTGGCCT AGGTGGGCTC TAGAACCAAT
2251 TACACAGCAT GAAAGAAACA GGTGGGTTA GGAGCAGAAA GAAATAAGGC
2301 TCACACCCCT CCAGACACTA CCTTATAAGC ACTGCAGAAC CTGAAACAGA
2351 TGGCAGAAGG AATGAATGC TACAGGGGCC AGCAGGAGTG ACCACAGGGA
2401 GGGGACAGCT CAGTGACTGG AGCATTACAG AAGAGGCTTT CCAGGGAACA
2451 CTGGACATTG CTTAGTGACC TTTTGTTCCT TTTTTTTTTT TTTTCTTTTA
2501 CTGTTCTGAA AGACTTTGAG TCTGTGGTTC ACCACCAGCC CATCAGTGTT
2551 TCTTTGAGGT GATTGCATTA GGAAGTTGG CTCTGGGATT GCAAAAAAAA
2601 AAAAAAGGTT GAACATGTTT TCCTTAAAG ATGGAAGGTT TTAGAAAATA
2651 TACTAGGCCA TCTGGTTAGA AAAAAAGAC CAGACTAGAA AAAGCTGTGA
```


2701 ATTTGATTTT GTAGATTAAA CAAAGCCAGA TGATTAAAT GTGATTTATT
 2751 TATAAAAAAA AAAAAAAAAA AAAA

BLAST Results

Entry HS463358 from database EMBL:
 human STS WI-14364.
 Length = 472
 Minus Strand HSPs:
 Score = 1605 (240.8 bits), Expect = 5.0e-68, P = 5.0e-68
 Identities = 347/361 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 16 bp to 960 bp; peptide length: 315
 Category: putative protein
 Classification: unset

1 MAAGVPCALV TSCSSVFSGD QLVQHTLGTE DLIVEVTSND AVRFPWTID
 51 NKYYADINL CVVPNKFLVT AEIAESVQAF VVYFDSTRKS GLDSVSSWLP
 101 LAKAWLPEVM ILVCDRVSED GINRQKAQEW SLKHGFELVE LSPEELPEED
 151 DDFPESTGVK RIVQALNANV WSNVVMKNDR NQGFSLLNSL TGTNHSIGSA
 201 DPCHPPEQPHL PAADSTESLS DHRGGASNTT DAQVDSIVDP MLDLOIQELA
 251 SLTTGGGDVE NFERPFSKLK EMKDKAATLP HEQRKVHAEK VAKAFWMAIG
 301 GDRDEIEGLS SDGEH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_46a6, frame 1

PIR:T04362 probable GTP-binding protein yptm3 - maize, N = 1, Score =
 87, P = 0.21

PIR:S71585 GTP-binding protein GB2 - Arabidopsis thaliana, N = 1, Score
 = 86, P = 0.27

>PIR:T04362 probable GTP-binding protein yptm3 - maize
 Length = 210

HSPs:

Score = 87 (13.1 bits), Expect = 2.4e-01, P = 2.1e-01
 Identities = 34/160 (21%), Positives = 67/160 (41%)

Query: 48 TIDNKYYADINLCVVPNKFL-VTAEIAESVQAFVVYFDSTRKSGLDVSSWLPLAKAWL 106
 TIDNK I F +T ++ +D TR+ + ++SWL A+
 Sbjct: 49 TIDNKPIKLQIWDTAGQESFRSITRSYYRGAAGALLVYDITRRETFNHLASWLEDARQHA 108
 Query: 107 PE---VMIL--VCDRVSEDGINRQKAQEWSLKHGFELVELSPEELPEEDDDFFPESTGVKR 161
 VM++ CD ++ ++ +++ +HG +E S + ++ F ++ G
 Sbjct: 109 NANMTVMLIGNKCDLSHRRAVSYEEGEQFAKEHGLVFMEASAKTAQNVEEAFIKTAGT-- 166
 Query: 162 IVQALNANVWSNVVMKNDRNQGFSLLNSLTGTNHSIGSADPC 203
 I + + ++ N G+++ NS G S A C
 Sbjct: 167 IYKKIQDGFIDVSNESNGIKVGYAVPNSSGGGAGSSSQAGGC 208

Pedant information for DKFZphfd2_46a6, frame 1

Report for DKFZphfd2_46a6.1

[LENGTH] 315

[MW] 34505.54
[pI] 4.55
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.67 %

```
SEQ MAAGVPCALVTSCTSSVFGDQLVQHTLGTEDLIVEVTSNDAVRFPWTIDNKYYRADINL
SEG .....
PRD cccccceeeccccccccccccccccccccccccccccccccccccccccccccccccce

SEQ CVVPNKFLVTAEIAESVQAFVVFYFDSTRKSGLDVSSWLPLAKAWLPEVMILVCDRVSED
SEG .....
PRD eeccccchhhhhhhhhhhheeecccccccccccccccccccccccccccccccccccccc

SEQ GINRQKAQEWSLKHGFELVELSPEELPEEDDDFPESTGVKRIVQALNANVWSNVVMKNDR
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cchhhhhhhhhccccceeeccccccccccccccccccccccccchhhhhhhccccceeecccc

SEQ NQGFSLNLSLTGTNHSIGSADPCHPEQPHLPAADSTESLSDHRGGASNTTDAQVDSIVDP
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ MLDLDIQELASLTGGGDVENFERPFPSKLEMKDKAATLPHEQRKVHAEKVAKAFWMAIG
SEG .....
PRD hhhhhhhhhhhccccccccccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhc

SEQ GDRDEIEGLSSDGEH
SEG .....
PRD ccccccccccccccc
```

(No Prosite data available for DKFZphfd2_46a6.1)

(No Pfam data available for DKFZphfd2_46a6.1)

DKFZphfkd2_46b10

group: kidney derived

DKFZphfkd2_46b10.1 encodes a novel 315 amino acid protein with similarity to C.elegans cosmide F25B5.3

The novel protein contains a HTH-LYSR-family PROSITE pattern. Proteins of the lysR family are bacterial transcriptional regulatory proteins which bind DNA using a helix-turn-helix motif. Most of these proteins are transcription activators and usually negatively regulate their own expression. They all possess a potential 'helix-turn-helix' DNA-binding motif in their N-terminal section. The 'helix-turn-helix' motif is missing in DKFZphfkd2_46a6.1. No informative BLAST results, no predictive PFAM or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to C.elegans F25B5.3

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 1285 bp

Poly A stretch at pos. 1266, no polyadenylation signal found

```

1  CAGTCTACGC GAGCTGCCTG TTTTTCCT GCTTGGACGC GCATGAGGGC
51  CCCGTCCATG GACCGCGCGG CCGTGGCGAG GGTGGCGCGG GTAGCGAGCG
101 CCAGCGTGTG CGCCCTGGTG GCGGGGGTGG TGCTGGCTCA GTACATATTC
151 ACCTTGAAGA GGAAGACGGG GCGGAAGACC AAGATCATCG AGATGATGCC
201 AGAATTCAG AAAAGTTCAG TTCGAATCAA GAACCTACA AGAGTAGAAG
251 AAATTATCTG TGGTCTTATC AAAGGAGGAG CTGCCAACT TCAGATAATA
301 ACGGACTTTG ATATGACACT CAGTAGATT TCATATAAAG GGAAGATG
351 CCAACATGT CATAATATCA TTGACAACTG TAAGCTGGTT ACGGATGAAT
401 GTAGAAAAAA GTTATTGCAA CTAAAGGAAA AATATTACGC TATTGAAGTT
451 GATCCTGTTC TTAGTGTAAG AGAGAAGTAC CCTTATATGG TGGATGGTA
501 TACTAAATCA CATGGTTTGC TTGTTCAAGC AGCTTTACCA AAAGCTAAAC
551 TTAAGAAAT TGTGGCAGAA TCTGACGTTA TGCTCAAAGA AGGATATGAG
601 AATTTCTTTG ATAAGCTCCA ACAACATAGC ATCCCGTGT TCATATTTTC
651 GGCTGGAATC GCGGATGTAC TAGAGGAAGT TATTCGTCAA GCTGGTGTTC
701 ATCATCCCAA TGTCAAAGTT GTGTCCAATT TTATGGATT TGATGAACT
751 GGGGTGCTCA AAGGATTAA AGGAGAACTA ATTCATGTAT TTAACAAACA
801 TGATGGTGCC TTGAGGAATA CAGAATATTT CAATCAACTA AAAGACAATA
851 GTAACATAAT TCTTCTGGGA GACTCCCAAG GAGACTTAAG AATGGCAGAT
901 GGAGTGGCCA ATGTTGAGCA CATTCTGAAA ATTGGATATC TAAATGATAG
951 AGTGGATGAG CTTTATAGAA AGTACATGGA CTCTTATGAT ATTGTTTAG
1001 TACAAGATGA ATCATTAGAA GTAGCCAAC CTATTTTACA GAAGATTCTA
1051 TAAACAAGCA TTCTCCAAGA AGACCTCTCT CCTGTGGGTG CAATTGAACT
1101 GTTCATCCGT TCATCTTGCT GAGAGACTTA TTTATAATAT ATCCTTACTC
1151 TCGAAGTGT CCCTTTGTAT AACTGAAGTA TTTTCAGATA TGGTGAATGC
1201 ATTGACTGGA AGCTCCTTTT CTCCACCTCT CTCACACAC TCCTCACCGT
1251 ATCTTTTAA CCATTAAAA AAAAAAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 43 bp to 1050 bp; peptide length: 336
 Category: similarity to unknown protein
 Classification: unset
 Prosite motifs: HTH_LYSR_FAMILY (16-47)


```

1 MRAPSMDDRAA VARVGAVASA SVCALVAGVV LAQYIFTLKR KTGRKTKIIE
51 MMPEFQKSSV RIKNPTRVEE IICGLIKGGA AKLQIITDFD MTLRSFSYKG
101 KRCPTCHNII DNCKLVTDEC RKKLLQLKEK YYAIEVDPVL TVEEKYPYMV
151 EWTKSHGLL VQQALPKAKL KEIVAESDVM LKEGYENFFD KLQQHSIPVF
201 IFSAGIGDVL EEVIRQAGVY HPNVKVVSNF MDEFETGVLK GFKGELIHVF
251 NKHDGALRNT EYFNQLKDNS NIILLGDSQG DLRMADGVAN VEHLKIGYL
301 NDRVDELLEK YMDSYDIVLV QDESLEVANS ILQKIL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_46b10, frame 1

SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III., N = 1, Score = 524, P = 2.2e-50

TREMBL:AC005499.12 gene: "T6A23.12"; Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence, complete sequence., N = 2, Score = 194, P = 1.4e-26

>SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.

Length = 376

HSPs:

Score = 524 (78.6 bits), Expect = 2.2e-50, P = 2.2e-50
Identities = 112/300 (37%), Positives = 174/300 (58%)

```

Query:   44 RRTKIIEMMPEFQ--KSSVRIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYK-G 100
      +KT ++ ++ + + ++ +PT V + ++ GGA K +I+DFD TLRSF+ + G
Sbjct:   73 KKTDDVPLLMNYLLGEEQILVADPTAVAAKLKRMVVGAGKTVVISDFDYTLRSFANEQG 132

Query:   101 KRCPTCHNIID-NCKLVTDEC RKKLLQLKEKYYAIEVDPVL TVEEKYPYMV EWTKSHGL 159
      +R T H + D N + E +K + LK KYI IE P LT+EEK P+M +W+ SH L
Sbjct:   133 ERLSTTHGVFDDNVMLRKLPELGQKFDLKNKYPIEFSPNLTMEEKIPHEKWWGTSHSL 192

Query:   160 LVQQALPKAKLKEIVAESDVM LKEGYENFFD KLQQHSIPVF IFSAGIGDVL EEVIRQA-G 218
      +V + K +++ V +S ++ K+G E+F + L H+IP+ IFSAGIG+++E ++Q G
Sbjct:   193 IVNEKFSKNTIEDFVRQSRIVFKDGAEDFIEALDAHNIPLVIFSAGIGNIIEYFLQQKLG 252

Query:   219 VYHPNVKVVS NFMDDETGV LKGFKEGELIHVFNKHDGAL-RNTEYFNQLKD NSNIILLGD 277
      N +SN + FDE F LIH F K+ + + T +F+ + N+ILLGD
Sbjct:   253 AIPRNTHFISNMILFDEDDNACAFSEPLIHTFCNSSVIQKETSFFHDIAGRVNVILLGD 312

Query:   278 SQGDLRMADGVANVEHLKIGYLNDRVDEL--LEKYMDSYDIVLVQDESLEVANSILQKI 335
      S GD+ M GV LK+GY N +D+ L+ Y + YDIVL+ D +L VA I+ I
Sbjct:   313 SMGDIHMDVGVERDGP TLKVGYYNSLDDTAALQHYEEVYDIVLIHDPTLNVAQKIVDII 372

```

Pedant information for DKF2phfd2_46b10, frame 1

Report for DKF2phfd2_46b10.1

```

[LENGTH]      336
[MW]           37948.37
[pI]           6.67
[HOMOL]        SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.
3e-51
[PROSITE]      HTH_LYSR_FAMILY 1
[KW]           TRANSMEMBRANE 2
[KW]           LOW_COMPLEXITY 7.44 %

```

```

SEQ  MRAPSMDDRAAVARVGAVASASVCALVAGVVLAQYIFTLKRKTGRKTKIEMMPEFQKSSV
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccchhhhhcchhhhhheehhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceehhhhhhhhhhe
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
SEQ  RIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYKGKRCPTCHNIIDNCKLVTDEC
SEG  .....
PRD  eccccchhhhhhhhhhhcccccceeeccccccccccccccccccccccccccccccccchhhhhh
MEM  .....

```



```
SEQ  RKKLLQLKEKYAIEVDPVLTVEEKYPYVWEYTKSHGLLVQQALPKAKLKEIVAESDVM
SEG  .....
PRD  hhhhhhhhhhhheeeccccccccccchhhhhhhccchhhhhccchhhhhhhhhhhcc
MEM  .....

SEQ  LKEGYENFFDKLQQHSIPVFIFISAGIGDVLEEVIHQAGVYHPNVKVVSNFMDFDGTGLK
SEG  .....
PRD  cccchhhhhhhhhccceeeeeeccccchhhhhhhhhccceeeeeeccccccccceee
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  GFKGELIHVFNKHDGALRNTEYFNQLKDNSNIILLGDSQGLRMADGVANVEHILKIGYL
SEG  .....
PRD  eccccccccccccccccchhhhhhhccceeeccccccccccccccccccccceeeec
MEM  .....

SEQ  NDRVDELLEKYMDSYDIVLVQDESLEVANSILQKIL
SEG  .....
PRD  cchhhhhhhhhhhheeeeeeccchhhhhhhhhcc
MEM  .....
```

Prosites for DKFZphfd2_46b10.1

PS00044 16->47 HTH_LYSR_FAMILY PDOC00043

(No Pfam data available for DKFZphfd2_46b10.1)

DKF2phfkd2_46d13

group: kidney derived

DKF2phfkd2_46d13 encodes a novel 506 amino acid protein with weak similarity to KE03 protein

The novel protein contains a RGD site.

No informative BLAST results; No predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KE03 protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="227.6 cR from top of Chr1 linkage group"

Insert length: 3346 bp

Poly A stretch at pos. 3328, polyadenylation signal at pos. 3308

```
1 CTCTCGCGAG AGGAGCAAGA GGAAGATGGC CGTGCCCTGT TTTTCGGTGT
51 AAGGCAGCAG ACGGCGGCTG CGACGGCGAG ACTGAGATCC TGGTGTCTGT
101 GGCACCTGAG TTCTAGCTTC CCCAGCGAG CGCGGTCCC TTCGTGCCTA
151 GCGAGAGACC GGCTCTTCCC CGGAGATGC GTTTGTCCCA GGCTCGGGGG
201 CTCACTGGGA GTTCATGCTG CGCTGGAGGC TCTTGGCCAC CGCTCTAATC
251 GCCTTGTGCC GCCGCAGCGC CAGCTCCGTC GCCAGCGGTG AGCCTCCCGA
301 TTCCCCCCTT TGCCCTTGGC GCGGCGGATG ACCGGGGAGA AGATCCGCTC
351 ACTGCGGAGG GACCACAAGC CCAGCAAGA AGAAGGGGAC CTGCTGGAGC
401 CCGGGGATGA AGAAGCGGCG GCTGCCCTCG GCGGTACCTT TACCAGAAGC
451 AGGATTGGCA AGGCGGCGAA AGCTTGTCTA AAGATCTTCA GTAACCATCA
501 CCACCGGCTA CAGCTGAAGG CAGCTCCGGC CTCCTCCAAT CCCCCCGGCG
551 CCGCGGCTCT GCGCTGCAC AATTCCTCCG TGAAGTCCAA CTCCTCAGTC
601 CCGGCGCTTC TGGCGGCGAC CAACCCCGTT GCTGTCTGTC CGGATGGAGG
651 CAGTTGCCCC GCACACTACC CGGTGCACGA GTGCGTCTTC AAGGGGGATG
701 TGAGGAGACT CTCTCTCTC ATCCGACGCG ACAATATCGG GCAGAAAGAT
751 AATCACGGAA ATACTCTCTT ACACCTTGCT GTGATGTTAG GAAATAAAGT
801 TACAGCTCTT TTGAGGAAGC TTAAGCAGCA ATCCAGGGAA AGTGTGAAG
851 AAAAACGACC TCGATTATTA AAAGCCCTGA AAGAGCTAGG TGACTTTTAT
901 CTAGCACTTC ACTGGGATTT TCAAAGCTGG GTGCCCTTAC TTTCCCGAAT
951 TCTGCCCTTC GATGCATGTA AAATATACAA ACAAGGTATC AATATCAGGC
1001 TTGACACAAC TCTCATAGAC TTTACTGACA TGAAGTGCCA ACGAGGGGAT
1051 CTAAGCTTCA TTTCAATGG GGATGCGGCG CCCTCTGAAT CTTTGTAGT
1101 ATTAGACAAT GAACAAAAG TTTATCAGCG AATACATCAT GAGGAATCAG
1151 AGATGGAAGC AGAAGAAGAG GTGGATATT TAATGAGCAG TGATATTTAC
1201 TCTGCAACTT TATCAACAAA ATCAATTTCT TTCACGCGTG CCCAGACAGG
1251 ATGGCTTTTT CGGGAAGATA AAACAGAAAG AGTAGGAAC TTTTGGCAG
1301 ACTTTTACCT GGTGAATGGA CTTGTTATAG AATCAAGGAA AAGAAGAGAA
1351 CATCTCAGTG AAGAGGATAT TCTTCGAAAT AAGGCCATCA TGGAGAGTTT
1401 GAGTAAAGGT GGAACATATA TGGAACAGAA TTTTGAAGCG ATTCGAAGAC
1451 AGTCTCTTAC ACCGCCTCCT CAGAACACTA TTACATGGGA AGAATATATA
1501 TCTGCTGAAA ATGGAAAAGC TCCTCATCTG GGTAGAGAAT TGGTGTGCAA
1551 AGAGAGTAAG AAAACGTTTA AAGCTACGAT AGCCATGAGC CAGGAATTTT
1601 CCTTAGGGAT AGAGTTATTA TTGAATGTT TAGAAGTAGT AGCTCCCTTC
1651 AAGCACTTTA ACAAGCTTAG AGAATTTGTT CAGATGAAGC TTCCTCCAGG
1701 CTTTCTGTGA AAATTAGATA TACCTGTGTT TCCCACAATC ACAGCCACTG
1751 TGACTTTTCA GGAGTTTCGA TACGATGAAT TTGATGGCTC CATCTTTACT
1801 ATACCTGATG ACTACAAGGA AGACCCAAGC CGTTTTCCTG ATCTTTAACT
1851 GACGTGGAAG AGGATGCCGT CTAACCAAGG AAAGAAAATA CAGAGACCCT
1901 AGAAGTGGAT CCAAATAGAA GGGACAAATG CTTTCAGTGA AGAAAAGGGA
1951 ATTACACATT GAATCGACAC ATCAGTAATA CGATACAGTG AAATGGGCCT
2001 CTAATAAGAA TTTCAGCGAG TTTTCTGATG TGCCATTTTT TGCTTTTTTA
2051 AAAATATACA TATTATAAAT GTAATAGTTT GACACATTAA TGACCCTAAG
2101 ACCTGCGTAT GTGAAGCAGC TATGAGTGCT GTGATTTGTT TTTAAAAATT
2151 TTTACACTTC TTGTTGAAAT ATATATGCAT ATAAATATAT CTATATCTAT
2201 ATCTATATCT AAAACACTCC TGGACCATTG ACGTAAATTA AATGCTTAA
2251 GAGATATGGA GCCCTTTTAA ACTTGTCTAT TTTATGCAAG GTGACATTTA
2301 TAAATATTTT TTCGAGCTTT GTTTTCATAA AATGTAAACT ATGTAACATT
2351 ATGTATAGTT CAGTAATTTG AATGTTTGT CAATATAATG AACTAGAAGG
2401 AATGCAATTT TCTGTAGATG AATGAACCAA ATGGTAACCA TTAACAATT
2451 GCATTTATAT GTTGCAATAC ATTTGAGAAG GAGCGTTTAC TCTGCAGGGA
2501 ATAAGGTACC TCCTTTAGCA CCTTAGTGCA ATTCATTGTG GTGCTATTTG
2551 TTTTACCTG AATGTTTGT ACTAATCTTC CTTTCATAGA ACCTCTATTT
2601 TTTTTTTTTT TAAACTTGAG TTTGAGTCTT TGTTATGGTC ATCATAAGGT
```



```

2651 AATGGTTAGC ATGTTTAAAG ATATTCCTCT TCCAAATCTC AGCACTTTAA
2701 AAAAAAATCC AAATTTTAA ACTTGCTTCC TAATAAGTAC ACATCGGTCT
2751 GATTATTTTG TTTGTTTTTA GTAGAATATG GATGCATTGG TGTCAGTTT
2801 AAAAAACAAT ACACATATTT TGGACAACCC TACATATTTA ATCCTTTCAA
2851 AATAAGATAA AAACATTTTA TATGCTAACA GAATATATTT GTTACAAGTT
2901 AAAGTCCAGA AGTATACACA AGATTGATTA CTCCTATTAT TTTTITAAA
2951 TCACAGGAAA ATATTGATTT CATTGTCTCC AAAGTGATAA AATCTTGTAT
3001 TACTCATTTC TGCACCTAAA ATTTTCTTA TTTATTCCAA GGTGGTTTGA
3051 AGGTCCAAGT ATGAAAATAA ATTAGGGGGA TTAATGTATA ACAGTTATAA
3101 AGTATCATGT TGTATTAAAG AGCTTACTTA GATTGATGTT TTTAAATGT
3151 ATCCTGATGA ATGTCTCAAG AATGCATCTG TCAAGTTTTT TAGACTGACC
3201 AGTAGCTTAA ACTTTTTTCA GGATTTTAGG TAATTGAAA GGAGTTTAGA
3251 GACCCTTATT GAAAATATGA TTTAAAAATC CAAAGCATAA ACCGTAAGAA
3301 AAATTTTAAA TAAACATCTT TAAAGCTGAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HS121353 from database EMBL:

human STS WI-14729.

Score = 1697, P = 1.9e-69, identities = 363/379

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 1845 bp; peptide length: 506
 Category: similarity to unknown protein

```

1 MTGEKIRSLR RDHKPSKEEG DLLEPGDEEA AAALGGTFTR SRIGKGGKAC
51 HKIFSNNHHR LQLKAAPASS NPPGAPALPL HNSSVTANSQ SPALLAGTNP
101 VAVVADGGSC PAHYVPVHECV FKGDVRRLLSS LIRTHNIGQK DNHGNTPLHL
151 AVMLGNKVTA LLRKLKQQR SRVVEEKRPRL LKALKELGDF YLELHWDFQS
201 WVPLLSRILP SDACKIYKQG INIRLDTTLI DFTDMKCQRG DLSFIFNGDA
251 APSESFVULD NEQKVYQRIH HEESEMETEE EVDILMSSDI YSATLSTKSI
301 SFTRAQTGWL FREDKTERVG NFLADFYLVN GLVIESRKRRL EHLSEEDILR
351 NKAIMESLSK GGNIMEQNFE PIRROSLTPP PQNTITWEEY ISAENGKAPH
401 LGRELVCRES KKTFFKATIAM SQEFPLGIEL LLNVLEVVPV FKHFNKLREF
451 VQMKLPPGFP VKLDIPVFPT ITATVTFQEF RYDEFDGSIF TIPDDYKEDP
501 SRFPDL

```

BLASTP hits

Entry CEC01F1_3 from database TREMBL:

gene: "C01F1.6"; Caenorhabditis elegans cosmid C01F1.

Score = 371, P = 4.5e-61, identities = 69/138, positives = 96/138

Entry CEC18F10_9 from database TREMBL:

gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10.

Score = 383, P = 3.4e-39, identities = 103/349, positives = 182/349

Entry AF064604_1 from database TREMBL:

product: "KE03 protein"; Homo sapiens KE03 protein mRNA, partial cds.

Score = 348, P = 8.3e-32, identities = 95/295, positives = 148/295

Alert BLASTP hits for DKFZphfkd2_46d13, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46d13, frame 1

Report for DKFZphfkd2_46d13.1

```

[LENGTH] 506
[MW] 57003.12
[pI] 6.40

```


(HOMOL) TREMBL:CEC18F10_9 gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10. 2e-35

(BLOCKS) BL01288E
 (PROSITE) RGD 1
 (PROSITE) MYRISTYL 7
 (PROSITE) CAMP_PHOSPHO_SITE 2
 (PROSITE) CK2_PHOSPHO_SITE 9
 (PROSITE) PKC_PHOSPHO_SITE 6
 (PROSITE) ASN_GLYCOSYLATION 1
 (KW) Alpha_Beta
 (KW) LOW_COMPLEXITY 7.51 %

```

SEQ  MTGEKIRSLRRDHKPSKEEGDLLEPGDEEAAAALGGTFTSRIGKGGKACHKIFSNNHHHR
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccccccchhhhhhhccccccccccccccccccccccccchhhhh

SEQ  LQLKAAPASSNPPGAPALPLHNSSVTANSQSPALLAGTNPVAVVADGGSCPAHYPVHECV
SEG  ....XXXXXXXXXXXXX.....
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  FKGDVRRLLSLIRTHNIGQKDNHGNTPLHLAVMLGNKV TALLRKLKQQSRESVEEKRPRL
SEG  .....
PRD  eccchhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhh

SEQ  LKALKELGDFYLELHWDFQSWVPLLSRILPSDACKIYKQGINIRLDTTLIDFTDMKCQRG
SEG  .....
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  DLSFIFNGDAAPSESFVVLNDEQKVYQRIHHEESEMETEEVDILMSSDIYSATLSTKSI
SEG  .....XXXXXXXXXXXXX.....
PRD  ceeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFTRAQTGWLFREDKTERVGNFLADFYLVNGLVIESRKREHLSEEDILRNKAIMESLSK
SEG  .....
PRD  eeeeeccccccccccccchhhhhhhheeeeeeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhc

SEQ  GGNIMEQNFEPIRRQSLTPPPQNTITWEEYISAENGKAPHLGRELVCESKKTFKATIAM
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  SQEFPLGLIELLLNVLEVAPFKHFNKLREFVQMKLPPGFVKLDIPVFPTITATVTFQEF
SEG  .....
PRD  hhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchc

SEQ  RYDEFDGSIFTIPDDYKEDPSRFPDL
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfd2_46d13.1

PS00001	82->86	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	373->377	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	336->339	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	278->282	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	386->390	CK2_PHOSPHO_SITE	PDOC00006
PS00006	476->480	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	46->52	MYRISTYL	PDOC00008
PS00008	108->114	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00016	239->242	RGD	PDOC00016

(No Pfam data available for DKFZphfd2_46d13.1)

DKFZphfkd2_46j20

group: metabolism

DKFZphfkd2_346j20 encodes a novel 224 amino acid protein similar to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein seems to be the human ortholog of 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

complete cDNA, complete cds, EST hits,
potential start at Bp 16 matches kozak consensus ANCatgG
strong similarity to proteins of worm plant archea and bacteria
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase is part of
the tyrosine metabolism (degradation of tyrosine late step) EC 5.3.1.-
complete cds according to similar C.elegans and A.thaliana protein

Sequenced by MediGenomix

Locus: unknown

Insert length: 1706 bp

Poly A stretch at pos. 1686, polyadenylation signal at pos. 1667

```
1  CACTTGATGG GAATCATGGC AGCATCCAGG CCATTGTCCC GCTTCTGGGA
51 GTGGGGAAAG AACATCGTCT GCGTGGGGAG GAACTACGCG GACCACGTCA
101 GGGAGATGCG CAGCGCGGTG TTGAGCGAGC CCGTGCTGTT CCTGAAGCCG
151 TCCACGGGCT ACGCGGCCGA GGGCTCGCCC ATCCTCATGC CCGCGTACAC
201 TCGCAACCTG CACCACGAGC TGGAGCTGGG CGTGGTGATG GGC AAGCGCT
251 GCCGCGCAGT CCCCAGGGCT GCGGCCATGG ACTACGTGGG CGGCTATGCC
301 CTGTGCCTGG ATATGACCGC CCGGGACGTG CAGGACGAGT GCAAGAAGAA
351 GGGGCTGCCC TGGACTCTGG CGAAGAGCTT CACGGCGTCC TGCCCGGTCA
401 GCGCGTTCGT GCCCAAGGAG AAGATCCCTG ACCCTCACA GCTGAAGCTC
451 TGGCTCAAGG TCAACGGCGA ACTCAGACAG GAGGGTGAGA CATCCTCCAT
501 GATTTTTTCC ATCCCTTACA TCATCAGCTA TGTTTCTAAG ATCATAACCT
551 TGGAAGAAGG AGATATTATC TTGACTGGGA CGCCAAAGGG AGTTGGACCG
601 GTTAAAGAAA ACGATGAGAT CGAGGCTGGC ATACACGGGC TGGTCAGTAT
651 GACATTTTAA GTGGAAAAGC CAGAATATTG AGTTATTCTT TAACAAGTTT
701 CGAGAGAGAA GGGAGCAAGA CAAGAGCAAG CAACGGCTAT TAAATGTCAC
751 AATCCTTTAA TTAGAAACCA TTTATTGGCC GGACGCGGTG GCTCACGCCT
801 GTAATCGCAG CACTTTGGGA GGCCGAGGCG GCGGCTCAC GACGTCAGGA
851 GATCCAGACC ATCTTGGCTA ACAGGGTGAA ACCCGTCTC TACTAAAAAT
901 ACAAAAAATT AGCCGGGCGT GGTGGCGGGC GCCTGTAGTC CCAGCTACTC
951 TGGAGGCTGA GGCAGGAGAA TCAATTGAAC CCGGGAGGCG GAGCTTACAG
1001 TGAGCTGAGA TTGCGCCACT GTACTCCTGG GCAACAGCGA GACTCCGTCT
1051 CAAAAAATAA AAAAAAATAA AGAAACCAT TATTTTAAAA ATGATTAGAT
1101 TGCTATGCCT CAATCATAG AAGATGAACC CTTCAAGAAA ACGTGAAGTA
1151 GAACGGGTGG GCCAGAAATG AAAACAGGCA AGTAAAGTAT TTCTTCGGAA
1201 AACATTTTAT CAAACCAAT GTTAAAAAGA CTTTCCTTTT GTAAAACTGG
1251 ATTAGAGAAG ACTTTTCAGT GGGTTATCTC TAGGATGATC AGTAGTTCAG
1301 CACTTAAAAA CTGCAGAGAA AACTGAAAGT TATGTTCCAG ATAACCTTCC
1351 GTTGTTTACC AAATTTTCTT AGATTGGTTC ATCATCAGGA AGCATTTGTA
1401 AAAATAAAAA TCTCCACAAA TTACTGGCCC ATCTCGGACT TGCTGAATCA
1451 ATTTGATAGG ATTAATCTCC AGTGAAGCTG TGTTCACAGG GCATTCCAAG
1501 TGATTCTTAT CAGGAAATGT GAAAAACACT CCGTACATA ATCGGTTAAT
1551 TTAAAAATTT ACTTAATAAG TGAACAAGTA ATGAAGATTT CACCTGTTTA
1601 CTTAGGGTAT CTACCCAGAC CCATCGATTG TGAGTTCGGG AGATGATTTT
1651 GAAATTACTG TTTTCCAAAT AAAGGTGCTC CCTTCCAAAA AAAAAAAAAA
1701 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

94039092: Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of *Escherichia coli* C.

Peptide information for frame 1

ORF from 7 bp to 678 bp; peptide length: 224
Category: strong similarity to known protein

1 MGIMAASRPL SRFWEWGKNI VCVGRNYADH VREMSAVLS EPVLFKLPST
51 AYAPEGSPIL MPAYTRNLHH ELELGVVMGK RCRAVPEAAA MDYVGGYALC
101 LDMTARDVQD ECKKKGLPWT LAKSFTASCP VSAFVPKEKI PDPHKLKLWL
151 KVNGLRQEG ETSSMIFSIP YIISYVSKII TLEEGDIILT GTPKGVGPVK
201 ENDEIEAGIH GLVSMTFKVE KPEY

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46j20, frame 1

PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*, N = 1, Score = 537, P = 8.7e-52

PIR:D71109 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase - *Pyrococcus horikoshii*, N = 1, Score = 529, P = 6.1e-51

PIR:C71425 hypothetical protein - *Arabidopsis thaliana*, N = 1, Score = 519, P = 7e-50

PIR:A64864 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase b1180 - *Escherichia coli*, N = 1, Score = 474, P = 4.1e-45

>PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*
Length = 214

HSPs:

Score = 537 (80.6 bits), Expect = 8.7e-52, P = 8.7e-52
Identities = 99/211 (46%), Positives = 138/211 (65%)

Query: 10 LSRFWEWGKNI VCVGRNYADH VREMSAVLS EPVLFKLPST AYAPEGSPIL MPAYTRNLH 69
L+ F IVCVGRNY DH E+ +A+ +P+LF+K ++ EG PI+ P +NLH
Sbjct: 4 LAGFRNLATK IVCVGRNYKDHALELGNAIPKKPMLFVKTVNSFIVEGEPIVAPPGCQNLH 63

Query: 70 HELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWT LAKSFTASC 129
E+ELGVV+ K+ + ++ AMDY+GGY + LDMTARD QDE KK G PW LAKSF SC
Sbjct: 64 QVELGVVISKKASRISKSDAMDYIGGYTVALDMTARDFQDEAKKAGAPWFLAKSFDGSC 123

Query: 130 PVSAFVPKEKIPDPHKLKLWLKVNGLRQEGETSSMIFSIPYIISYVSKII TLEEGDIIL 189
P+ F+P IP+PH ++L+ K+NG+ +Q T MIF IP ++ Y ++ TLE GD++L
Sbjct: 124 PIGGFLPVSDIPNPHDVELFCKINGKDQQRCTDVMIFDIPTLLEYTTQFFTLEVGDVVL 183

Query: 190 TGTPKGVGPVKENDEIEAGIHGLVSMTFKVE 220
TGTP GV + D IE G+ ++ F V+
Sbjct: 184 TGTPAGVTKINSGDVIEFGLTDKLNKFNQ 214

Pedant information for DKFZphfkd2_46j20, frame 1

Report for DKFZphfkd2_46j20.1

[LENGTH] 224
[MW] 24843.07
[PI] 6.96
[HOMOL] PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans* 8e-55
[FUNCAT] r general function prediction [M. jannaschii, MJ1656] 9e-40
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YNL168c] 4e-38
[EC] 5.3.3.10 5-Carboxymethyl-2-hydroxyruconate delta-isomerase 1e-35
[PIRKW] isomerase 1e-35
[PIRKW] intramolecular oxidoreductase 1e-35
[SUPFAM] 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase 1e-46
[PROSITE] MYRISTYL 4
[PROSITE] AMIDATION 1

{PROSITE}	CK2_PHOSPHO_SITE	2
{PROSITE}	PKC_PHOSPHO_SITE	3
{KW}	Alpha_Beta	

```

SEQ  MGIMAASRPLSRFEWGNIVCVGRNYADHVREMRSAVLSEPVFLKPSTAYAPEGSPIL
PRD  cccccccccchhhhhccceeeecchhhhhhhhhccccccceeecccccccccccccc

SEQ  MPAYTRNLHHELELGVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWT
PRD  cccccchhhhhhheeeccccccccchhhhhhhheeeecchhhhhhhhhhhcccccc

SEQ  LAKSFTASCPVSAFVPKEKIPDPHKLKLWLKVNDELROEGETSSMIFSIPYIISYVSKII
PRD  cccccccccceeeccccccccceeeccccccccccccceechhhhhhhhhhh

SEQ  TLEEGDIILTGTGPKGVGPVKNDEIEAGIHGLVSMTFKVEKPEY
PRD  hccccceeeccccccccccccceeecccccccccccccc

```

Prosites for DKFZphfd2_46j20.1

PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	192->195	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00008	2->8	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00009	78->82	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfd2_46j20.1)

DKFZphfkd2_46k19

group: transcription factors

DKFZphfkd2_46k19.3 encodes a novel 130 amino acid protein similar to rat Dcoh, a bifunctional protein-binding transcriptional co-activator.

Dcoh is a bifunctional protein, complexed with biopterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the biopterin cofactor of phenylalanine hydroxylase.

The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

strong similarity to pterin-4-alpha-carbinolamine dehydratase

potential start at Bp 102 according to similar proteins,
both genomic sequences are from chromosome 5,

Sequenced by MediGenomix

Locus: map="5"

Insert length: 5641 bp

Poly A stretch at pos. 5617, polyadenylation signal at pos. 5598

```
1 CAGCCCTCGG CAGACGGCCA ATGGCGGCGG TGCTCGGGGC GCTCGGGGCG
51 ACGCCGCGCT TGTTGGCGGC GCTGCGAGGC CAGAGCCTAG GGCTAGCGGC
101 CATGTCATCA GGTACTCACA GGTTGATTGC AGAGGAGAGG AACCAAGCTA
151 TACTTGACCT TAAAGCAGCA GGATGGTCGG AATTAAGTGA GAGAGATGCC
201 ATCTACAAAG AATTCTCCTT CCACAATTTT AATCAGGCAT TTGGCTTTAT
251 GTCCCGAGTT GCCCTACAAG CAGAGAAGAT GAATCATCAC CCAGAAATGGT
301 TCAATGTATA CAACAAGGTC CAGATAACTC TCACCTCACA TGAATGTGGT
351 GAACTGACCA AAAAAGATGT GAAGCTGGCC AAGTTTATTG AAAAAGCAGC
401 TGCTTCTGTG TGATTCTTTC CAAAATACAT AAGTCTGAGA GGCTAAACTT
451 GATGGCTGTG TTAACATATG TCACGTGTAG CACAGTGGAG AAAGCAGGAT
501 ATGGCTCATA ATGACAGTGG TGAAGACCTG CGAATGAAGT TGCTAGTTAA
551 CACCTACATT AGGGTTTGAC ATAGGTCTAT GTTATGGGTC GCTGCATCTG
601 CTGGAACCTCA CAGACTTTAC TATAGAGAAT CAAAGATCCC GTATCCGAAG
651 TCTATGGAAA TGCTCATGGT GGTAAATTCC AACAGAATGA AACACCAAAC
701 TTGCTTAAAG TAACTCACGT TTCAATTGGA AAGAGATATT GTCAAAATTG
751 GAGGCCCCCA GGTTCCTGTC TGTTCCAAAT CTTTGCATGA TGACAGTGGT
801 TTCTCTGATG TGGTAAGCTT TGGCTTTCTT CTGTTTCTT TCTAAAAGAT
851 CACTGGAGTA GAGAGGAGTT AAACAGACAT GACCTTTGAC CTCTTGATG
901 ACCTCCACAG ATAGCAAACC GGGCCGACAC ATGGTTGACG ATGTCCTTTT
951 CTACAATGAA GTTAATGAAA GTTCTGAAAA TAGTGATTAC TTTCTGACAT
1001 TGATAGGATT TAGGAAACCT CTGGATAAAT AGCTTAAGCA TGGCTGTTTA
1051 TGTTTTTGCT ATAGACAAAA AGCAGCAGCA TGTACATTGT ATTTGGACAC
1101 AAGCCTGCCT CGGTAAATAT ATTGAACAT TGGACCACTA GGGTTAGTAG
1151 GGAGCGGTCT GTACACTTTC TGATTCAGCA TTCAGAAACA TTCTAGGTGG
1201 ACTCTGTAGC TTTCAGTTT GTAAAGTTAT CGGAAAAACA TCGGGAGGGT
1251 TTGGCCATCA TATGTGAGCT TTGTGTTTCA ATGCCAGTTA CTCAGGATTA
1301 GTAAATTAAT GACTGTCCAG AGGACTTCAG GGTCAACCAAG CTGCTGCACC
1351 TGCCATTGGC TGAATCTCCC CGGCTATCTG TGGCTGAGAT GGTGCTGCTT
1401 AGGTCACGCA GAGCATGAGC TGCTGCTGAA AGGACACAGG AGATGGCCCT
1451 TGGGCTTCTC ATCCCAGGAT GCCTGCCCTG CCCACCAATC CATGAGAAGA
1501 TATGTATGAT TTCAGTAGGC CCTGGATCAG CTTGTCACCT CTGGTTTCTT
1551 GTTTGCTTTC CACTCACTCA GCTGGAGTTT CATTTCAGCA CTAAAGTCTT
1601 CATCATTTGGC TTCAGAAACA GCATTCACTT GTGGCTGTGC TGATGTAGTA
1651 CACCAAGAAC AACTGGGCTC TTCTCTGTCA CTTTCAGTGG GCTACCTTCC
1701 CTCACCTCTC CAAGCAGCAT GAAAGAATTC TTTACATTTT TAATCTCTTT
1751 TTTGTTTTTC CTGAAAGTA TGCTTTGGTG CTTAAAGAGA GAAGTCACAA
1801 AAGTATACTA CTGAGTTTCC TGGAGATGAA ATCCTGTTGT CCCTAGCTAT
1851 GTGAATGAGC ACAGGGATCC CTGATGCCAT TATTTGTAT ATTCATACGG
1901 CACACACTTA CTGAGGGCCT TCTGTGTGCC CTAGGGGATT GAGCACAGTG
1951 ACATATCAGG GCAGGTAGAA ACAGATGGAG AGCTGATGCG GGCTGCTCTA
2001 GAGCAGCTGC CCCAGGAGGC CCCTGTGGAT GGATGTTGGG CAGGAGCCCT
2051 GAGACGTTAG GGGCATATAA CTAAAGGACA TAGCAGGAGT TATAGGAGGA
2101 GCTGATCCCT GAGGAAACA ATGAAGACGG AGAAGATGGG GCTAAAGTTT
2151 GAATTGTGGG GACATTAATC ACGGTGATTC TTAACACTTT GCTGTTGATG
2201 ATTTTAAATG GAGAAAATGA GTACGTAAGA TGTTATTTC CAGTTCAGTA
2251 TATAGGTTGC CCACAAAGTA TTTTCTTACC ATGAATGGTC ATATATACTT
2301 GTTGTAGAAT ACCAGGGACA GCAGAGATGG TGGGGTAGTT ACTTCCTTTT
2351 CTTACAGCCC AAGAACTTTG GTGTCCAGGA GATTGACCAA TTTAGCCACT
2401 GAGCATTTAA TACAACACAG GGCTACCCAG ATCCCACTGT CCGATTGTC
2451 CCTGAAAGCC AAGGAGTCA GGAGAAGGTG AGTGGGGTGA ATATATTAAT
2501 CCTGAGAGTT GAACAGAGCA AAAATCCCTA TTACTTTTGT ACTTAAACA
```



```

2551 TCTCTGCCAC ATGTGCTCAC TCTTTATATT CTGTTTAGGT GGTATTATATG
2601 TGCACATCCC ATCCTATGCC TGCAGTTAGC CAACTCAGGG TTTATATTGC
2651 CTCCTTTCTT TTTTCTTTT TTTTCTTTT TTTTAAGAGA TGGGGTCTCG
2701 TTCTGTCTAG CAGACTGGAG TGCAGTGGTG TGATCACAGC TCATTGTAAC
2751 CTCCAACGCC TGGACTGAAG TGATCCTCCT GCCTTGGCCT CTCTGGTAGC
2801 TGGGACTACA GGTGCATGCC ACCACACCCA CCTAATTTT TTTATTTTAA
2851 TTTTGTGTAG AGACAGTCTC ACTATCTTGC TCGGGCTGGT CCTGAACCTCC
2901 TGGGCTCAAG TTATCTTGCT GCCTCAGCCT CCCATGGGTA ATCTTTATTT
2951 CCTTTTCTT TTTTCTTTG AGATGGAGTT TCGCTCTTGT CGCCAGGCT
3001 GGAGTGCAT GGCACGATCT TGGCTCACTG CAGTCTCCAC CTCTGGGTT
3051 CAGGTGATTC TCCATCCTCG GCCTACTGAG TAGCTGAGAT TACAGGCAAC
3101 TGCCACCATG CGCGGCTAAT TTGTGTATTT TTTTATAGTA AGAGATGGGG
3151 TTTCCGCATG TTGGCCGGAC TGGTCTTAGA CTCTGACCT CAAGCGACCT
3201 GCCTGCCTTG GCCTCCCAA GTGCTGGGAT TACAGGCATG AGCCGCTATG
3251 CCTCGTCGCT GATTTTATT TCTTATTTT TTTTATAGTA TGGGGTCTC
3301 ACTATGCTGC TCAGGCTGAT CTCAACTCC TGGCCTCAAG TGATCCTCCC
3351 ACCTTAGCCT CCCAAGTTGC TGGGATTATA AGTGTGAGCC ACTATCCCTA
3401 CCTCACTATT ACCTTCTTTG CTTCTCTTGT TTTCTTTTGT TCTAAGTCAA
3451 ACCCATCACA ATCTTTCTT GTCTTCCAG GTGTTTCCA GTGCTGTGCC
3501 CTGGATGTGC TCTCTTCTC TTAGAGCCCA GAGAAGTTGC TTTTCCCTT
3551 TATATATGAC CCTTAACCTT TTCTAACACA TTATTAAGGG CCTGTGTCTA
3601 TCAGCTGGGG GCACCTCTTG AAGGGAGGGC CTTTGTGTGG TCTGTTTCTA
3651 GTGACTTCCA GCTTTAACC AGAGCCTCAT GATTGCTGGG TGCCCATAGC
3701 CTTTTTGCTG AATGGAGGCA CTCAGTCTCC TTGGGAAGAG AGAATCCATG
3751 ATAGACCCAC TTGGGAGCTC CCCACTTCAG GGGCCTACAC ACTGGTAATG
3801 CAACAGAATG CCCAAGAGTG ACCTCATAAA GCAAGGATTC CCTTCGTGGC
3851 CCCTTCTCTG CTGCCTCTCA GAATCCAGAC GCTAAGGAAA ATCCCTAAGC
3901 AGAGATTTTC TGTGGATGTC TAAAAGCAAG GAATAAAAGT TGAAAATTTG
3951 GAAAATGTCT CAACACCGTC ACCAGCGCCA CTCGAGAGTC ATTTCTAGTT
4001 CACCAGTTGA CACTACATCG GTGGGATTTT GCCCAACATT CAAGAAATTT
4051 AAGTAAATAT TATCTATCTC CATTGCCTGT TAAGAAATGT GCTAGTAGAA
4101 GTGTGAGGGC AGGGTGTCTG TGTCTCTCA GCCTCTTCCC TCAGATACTC
4151 GTCTGCTTAC CAAAATAAGT TGATGTCTCT TGACATCTG GTTCTATGA
4201 TTGGTGAGGC TGGCATGCTA TTACCTTTAT GTGCCCTGTA GACTTGAATG
4251 ACCAGTTTGA CCAGTTTGAC TGTTAGATAA TCAGAAGGCT TTTCTCTTTT
4301 TTTATAATAG ACCCATCTC AAATCAGATA ATGAAAATTA CATATCTTGA
4351 TATATTAGAA AAGTATATAC ATTCTGGCTG GGCACGGTGG CTCACGCTG
4401 TAATCCCTGC ACTTTGAGAG GCTGGGGCGG ATCACTTGAG GTCAGGAGTT
4451 TGAGACCGGC CTGGCCAGCG TGGCGAAACC CCATCTCTAC TAAAAATACA
4501 CAGATTAGCC CGGAGTGATG GTGTGCACCT GTTGTCCCAG CTAATCAGGA
4551 TGCTGAGGCA GGAGAATCCC TTTAACCCTGG GGGGCCAAGG TTGAGTGAG
4601 CCAGGATTGC ACCACTGCAC TCCAGCCTGG GTGACGGAAC GGGACTCTGT
4651 CTCAGAAAAA AAAAAAAGA AGAGGAAAAA GAAAAATATA TATTCTATAT
4701 TTTTAACTAT TATGAGAATG TGTTCAATTC ATTTGTAACA TATAATGGGA
4751 AACAGTAATA CGTACTCTGA GAAAAATTGC AAAGCACAGA TAAATGGAAA
4801 TAAACAGGAA AAAGAATCAC CTATAACCTC ACCATCCATA GACAGACACT
4851 GTTAAATTTT TGGCATATTT CCTGCTGATT TTTTCTACTG CTGATTTTGT
4901 CACAGGTGAG ATAATTTTGA ACAGAGAATT TTGTATCTTT GGTTTTGTG
4951 TTTCCGCTGA CAAAAACA AAAGATATAA AAATGGATCA TAAACATTTT
5001 TCTAAATCCT GAAAAGTGCA TAGACATATT TTAGTGCTTG TATTTCACAA
5051 GATGGACATA CCATAATTTA CTTACACAGT CCTTTTGTG AGATGTTTAA
5101 GTTGTTTTCA AGCTTCTCAG TGCTGGAAAA AACTGTGAGA TAGACATGTT
5151 TAGTTGAAGT TATTTCAAT CAGGTTATAT TATCTTGGGT CAGAGAATGA
5201 ATGGTTCTCA GGCTTTTCAA AAGAGCTGGT CAGTTTTTAT GCCTCTGGCA
5251 GTTTTGTAGA GTGCTCAATC ATACTACACT GTTGCCAGCA TTAGATCTTA
5301 TCACATTTAA GTCATTGCTA ATTTTATAAA CAAAAACAAT GGTTTTACTT
5351 TGCATCTCCC TGATTGGTGT TGCTGTAGAA CATATTTGGA GAAGTTTGTG
5401 TGTCTTTTGT GTTTATTCCA TGAATAGATT GTGTGCCCAT TTTCTCTTGG
5451 GGTATTCACT TTTTATTAC TGATGTGAGC ATGTGTATGG GTGATTATTT
5501 GATGATTATC AGTTTTGCTT AGTAGACTGG CAATATTTAG TCTTGCTGTC
5551 ACTGTGTTCC CAGTGCCAAC TAGATTGCTT GATATGTAGT TGCCACTCAA
5601 TAAAGATTG TTAGTCAAT GAAAAAATA AAAAAAATA A

```

BLAST Results

Entry AC004764 from database EMBL:
Homo sapiens chromosome 5, P1 clone 255g5 (LBNL H61), complete
sequence.

Score = 11057, P = 0.0e+00, identities = 2217/2224
Bp 428-5625 of cDNA == Bp 2912-8107 of AC004764

Entry HSAC1555 from database EMBL:
Homo sapiens (subclone 1_d8 from BAC H75) DNA sequence, complete
sequence.

Score = 575, P = 5.1e-30, identities = 115/115
Bp -240- 430 of cDNA == HSAC1555 splice pattern

Medline entries

93186787:
Phenylalanine hydroxylase-stimulating protein/pterin-4
alpha-carbinolamine dehydratase from rat and human liver.
Purification, characterization, and complete amino acid
sequence.

93101632:
Identity of 4a-carbinolamine dehydratase, a component of
the phenylalanine hydroxylation system, and DCoH, a
transregulator of homeodomain proteins.

95242099:
Crystal structure of DCoH, a bifunctional, protein-binding
transcriptional coactivator

Peptide information for frame 3

ORF from 21 bp to 410 bp; peptide length: 130
Category: strong similarity to known protein

1 MAAVLGALGA TRRLAALRG QSLGLAAMSS GTHRLIAEER NQAILDLKAA
51 GWSELSEDA IYKEFSFHNF NQAFGMSRV ALQAEKMNH PEFWNVYKVV
101 QITLTSHDCG ELTKKDVKLA KFIEKAAASV

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46k19, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46k19, frame 3

Report for DKFZphfkd2_46k19.3

```

[LENGTH]      130
[MW]           14377.56
[pI]           9.17
[HOMOL]        PIR:A47189 pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) - rat 4e-34

[FUNCAT]       01.07.99 other vitamin, cofactor, and prosthetic group activities (S.
cerevisiae, YHL018w) 5e-04
[SCOP]         dldchg_4.38.1.1.1 Pterin-4a-carbinolamine dehydratas 4e-50
[EC]           4.2.1.96 Tetrahydrobiopterin dehydratase 6e-34
[PIRKW]        nucleus 6e-34
[PIRKW]        carbon-oxygen lyase 6e-34
[PIRKW]        homotetramer 6e-34
[PIRKW]        hydro-lyase 6e-34
[PIRKW]        cytosol 6e-34
[PIRKW]        acetylated amino end 6e-34
[PIRKW]        homodimer 6e-34
[SUPFAM]        pterin-4-alpha-carbinolamine dehydratase 6e-34
[PROSITE]       MYRISTYL      2
[PROSITE]       CK2_PHOSPHO_SITE      3
[PROSITE]       PKC_PHOSPHO_SITE      4
[KW]            Alpha_Beta
[KW]            3D
[KW]            LOW COMPLEXITY      14.62 %

```

```

SEQ      MAAVLGALGATRLLAALRGQSLGLAAMSSGTHRLIAEERNQAIIDLKAAAGWSELSEKDA
SEG      .XXXXXXXXXXXXXXXXXXXXX.....
ldchB    .....CCCCHHHHHHHHHHHHHHHCCCEEECCCE

```

```

SEQ      IYKEFSFHNFNQAFGMSRVALQAEKMNHPPEWFNVYNKVQITLTS HDCGELTKKDKVLA
SEG      EEEEECCCHHHHHHHHHHHHHHHHHHHHHCCCCCEEETTTEEEEECBTTTTBTCCHHHHHH
ldchB

```

```

SEQ      KFIKAAASV

```


SEG
ldchB HHHHHHHHHH

Prosites for DKFZphfkd2_46k19.3

PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	32->35	PKC_PHOSPHO_SITE	PDOC00005
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113->117	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_46k19.3)

DKFZphfk2_46m4

group: signal transduction

DKFZphfk2_46m4.3 encodes a novel 198 amino acid putative GTP-binding protein related to the SAR-1 family of Ras superfamily members.

SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

The new protein can find clinical application in modulating the transport of vesicles to the Golgi Apparatus, thus enabling post-translational modifications of the vesicles contents.. Blocking of the molecule is expected to result modulation/blocking of secretory pathways.

nearly identical to mouse GTP-binding protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="438.9 cR from top of Chr10 linkage group"

Insert length: 2996 bp

Poly A stretch at pos. 2969, polyadenylation signal at pos. 2958

```
1 ACATCCGGCG AGTAGCTGGC GGTCCCGGGT GCTGCTGGTT AGTGTGCTCT
51 GAGGGAGGGT CCGAGCCAGC CGCTGTTTTG CCGGAGGAGC CCCTCAGGCC
101 GTAGTAAGCA TTAATAATGT CTTTCATCTT TGAGTGGATC TACAATGGCT
151 TCAGCAGTGT GCTCCAGTTC CTAGGACTGT ACAAGAAATC TGGAAAACCT
201 GTATTCTTAG GTTTGGATAA TGCAGGCAAA ACCACTCTTC TTCACATGCT
251 CAAAGATGAC AGATTGGGCC AACATGTTCC AACACTACAT CCGACATCAG
301 AAGAGCTAAC AATTGCTGGA ATGACCTTTA CAACCTTTGA TCTTGGTGGG
351 CACGAGCAAG CACGTCGCGT TTGGAAAAAT TATCTCCCAG CAATTAATGG
401 GATTGTCTTT CTGGTGGACT GTGCAGATCA TTCTCGCCTC GTGGAATCCA
451 AAGTTGAGCT TAATGCTTTA ATGACTGATG AAACAATATC CAATGTGCCA
501 ATCCTTATCT TGGGTAACAA AATTGACAGA ACAGATGCAA TCAGTGAAGA
551 AAAACTCCGT GAGATATTTG GCCTTTATGG ACAGACCACA GGAAAGGGGA
601 ATGTGACCCG GAAGGAGCTG AATGCTCGCC CCATGGAAGT GTTCATGTGC
651 AGTGTGCTCA AGAGGCAAGG TTACGGCGAG GGTTCGCT GGCTCTCCCA
701 GTATATTGAC TGATGTTTGG ACGGTGAAAA TAAAGAGTT TTAATCTCT
751 GGACTGATCC TATTCACAGC TTCCTCATGA ACTTTCTTAA TAGAACAAGG
801 ATAGCTCTCC AACCATGTCT GCGGTTGAGA AGCCAAGAGT CTCTGTCAAC
851 TCTCTCATTT CCCAGTGGTG ACATGTGCTC TTCTCCACAC TGTGGGAGG
901 TAATGCTGCC CCACGTGCTG GTGCAGGTCA GTATCCTGGG ACTTGGGAAGC
951 TGGCAGGATT TGCCGGGTAA AGCTGTATGC CATCATGGGG CACCTGAAAA
1001 GAAAAACACG TCTCACCACG GTGGTTGATT CAAAAGAAAG TGATTCTATT
1051 TTTTAAAGAA AGCGTTGTTA ATGTAATTGG TATCCCTCCT AACTTTTTGA
1101 GTTCACAAAT TACTTGGTCC AGAGTTTTCT ATTCTTTTTT TTTTTTAAA
1151 CTAATGAATG ACATTTAGAT ACTTCATAAA ATTATGAACA GATATGGAGG
1201 CCAGAGCTCA TTTGGGTAAA CTTACTCCTG CTGAGTTAGC AGGTTGGTGA
1251 GAGAAGCTCC CCTGAGCTCA CTTGCTCTCT TGAAGCTTGG GAGTAGGTG
1301 GCATAACCTT GTGCACAGAG AACTAGAAAA GGGGCAGAAC CCCGGCCTTG
1351 CAGTTGTGGC AGGTTTCCAC TGTGGTAAGC TAGGTTTATT CCTCATCAAG
1401 GAATGTGTAG CAGATTGTTT ACTGTGGAGG AGGTAATTAT AGAATGGGTT
1451 ATTGTGTGTT TTCTTACTCA TGAAGTTACA GATTTTAGCC AGTCTTTGCT
1501 TTTTACTTTT TGTGAAATTT AATTTCTCTC TATAGCACCT TCCTTTTTTC
1551 TTTTCAGTTA TCAAAAGTGA CTTTGACCTC ATAAGAGAGT TGAGAACATC
1601 TCTCGTGTCA CATACTGCAG GTGCATCAGT TACTTTTGCA CAGATTCTAG
1651 GGGGACATTT TTCTGAATAG GAAGACAGGA CAAAGTTAAC AGCTTAAGGG
1701 CTCTTAATTC TGTGAGTTGA GGACTTAAAA GTATTGTAGC ATTTGTTTGG
1751 ATCCATGAAA AATGTATTCA GTGGGCTTTA AAATTTCCAT TTGCAGAATT
1801 TGGTCTCTCA GGCTGTTTGG GAGCTCTTTT TTTTACATTT TTTCTCCTTT
1851 GACACCTATT TTATTGGTGT TTAAAGTAAA GGTAAACATC TGTAGCTTTT
1901 CCAGGTTTTT TTTTTTTTTT TTGATATGAA ATTGTCTTTC TCCATTGCAG
1951 AAATAAGCTA GGGAAACACT AACCCAAAAA CTTCTGTAG AGCTGTTCCCT
2001 TTGGAGGCAG CATCACTTAT TGGCAGTAAA GACTCAGTAT AAAAGCACCA
2051 GCATCCCTAC TTGGGTGATG GGGATTAAAT TTATAGCATT CCATTTTCCT
2101 AGTGCCACAT GTGAAATTGG ATTTTGATGA TCTTAATCTA TATTCTACCC
2151 TTATAATAAA AGATCAAAAG ATATATCTCC TATGAACAGA TTGGAGATAG
2201 GTGATGAAAA GTTGGGAGGA TGCCTTTATT CTAATGTGAG GGTAGGGAAA
2251 ATGTGGATAA CATTACTGGG GTGAAGGAGG CATGTGCTT TAGTTGGAGT
2301 TCTCATTTTT ATTCTCCAGT ACTGACTTGT GGGGAAAGCA TACTTTTTCA
2351 CTGCCAGGTA CTGAATGCAG AGGCTCAGTG AAGTATATAT GTGGGAAGTG
2401 CATGCATTTT GTTTATTAGC AAACATAGCT GGATTAAGAC GAAGTTGTTG
2451 GTTTGGAAAG GGGTTAAAGC CTTAAGTGAA CAAATCTAGC TAACAGTGAA
2501 TGAACTAGGT AATATAACTT GCATATTTTT AATTTCTTTT GGTAAAGGT
2551 CCCCCATACT TCTCTGTTTC GAGACATGAG AAGTATGATT ACTTCAGTGT
```



```

2601 TAGTTTCTT AATTTTTTTT TTCCCTATT TGTCCCTGT CACTTGTG
2651 CAAGCTAGAA ATCTGTGGGT TATACATAGG GCAGCTCTT GCGAAAGTGG
2701 TTTATTCCAC TGGAGAAAGG GGATTGAAAA TCAGTTAGAA CCAATGTATT
2751 TCTTGCCCA CGGAACACTA TTCCTATAAG ATAGCTGAAA GAAGCTGCTG
2801 TGAGGAGCTC AGCTCCAACA CAGGATCAGC ACCTTGATA GGAATTCCTA
2851 TGAATTATGA CTTCTCATT TGTTTATCA GAGTGCATAT ATGTCCTACT
2901 TCAGGAAAAG TAAACAGTC ATTTACGAAA GAAAGTCAAT CTGTATCCTA
2951 AGCATTTTAA TAAAAAGTTA AAACAAAAAA AAAAAAAA AAAAAA

```

BLAST Results

```

-----
Entry HS679348 from database EMBL:
human STS WI-16722.
Length = 265
Minus Strand HSPs:
Score = 1242 (186.4 bits), Expect = 2.8e-50, P = 2.8e-50
Identities = 260/265 (98%)

```

Medline entries

```

-----
94085558:
Molecular analysis of SAR1-related cDNAs from a mouse
pituitary cell line.

```

Peptide information for frame 3

```

-----
ORF from 117 bp to 710 bp; peptide length: 198
Category: strong similarity to known protein

```

```

1 MSFIFEWIYN GFSSVLQFLG LYKKSGLVF LGLDNAGKTT LLHMLKDDRL
51 QGHVPTLHPT SEELTIAGMT FTFDLGGHE QARRVWKNYL PAINGIVFLV
101 DCADHSRLVE SKVELNALMT DETISNPIL ILGNKIDRTD AISEEKLREI
151 FGLYGQTTGK GNVTLKELNA RPMEVFMCV LKRQGYGEGF RWLSQYID

```

BLASTP hits

```

Entry S39543 from database PIR:
GTP-binding protein - mouse
Length = 198
Score = 1029 (362.2 bits), Expect = 5.1e-104, P = 5.1e-104
Identities = 197/198 (99%), Positives = 198/198 (100%)

```

```

Entry SARA_MOUSE from database SWISSPROT:
GTP-BINDING PROTEIN SARA.
Length = 198
Score = 1012 (356.2 bits), Expect = 3.2e-102, P = 3.2e-102
Identities = 195/198 (98%), Positives = 196/198 (98%)

```

```

Entry CE2K180_4 from database TREMBL:
gene: "ZK180.4"; Caenorhabditis elegans cosmid ZK180.
Length = 193
Score = 679 (239.0 bits), Expect = 6.3e-67, P = 6.3e-67
Identities = 125/197 (63%), Positives = 161/197 (81%)

```

Alert BLASTP hits for DKFZphfkd2_46m4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46m4, frame 3

Report for DKFZphfkd2_46m4.3

```

[LENGTH]      198
[MW]           22367.00
[pI]           6.21
[HOMOL]        PIR:S39543 GTP-binding protein - mouse 1e-112

```


[FUNCAT] 08.07 vesicular transport (golgi network, etc.) (S. cerevisiae, YPL218w)
 1e-58
 [FUNCAT] 30.09 organization of intracellular transport vesicles (S. cerevisiae,
 YPL218w) 1e-58
 [FUNCAT] 06.10 assembly of protein complexes (S. cerevisiae, YOR094w) 2e-23
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) (S. cerevisiae, YPL051w) 4e-22
 [FUNCAT] 30.08 organization of golgi (S. cerevisiae, YDL192w) 3e-20
 [FUNCAT] 30.03 organization of cytoplasm (S. cerevisiae, YBR164c) 3e-19
 [FUNCAT] 03.22 cell cycle control and mitosis (S. cerevisiae, YMR138w) 2e-09
 [FUNCAT] 30.04 organization of cytoskeleton (S. cerevisiae, YMR138w) 2e-09
 [FUNCAT] 98 classification not yet clear-cut (S. cerevisiae, YHR168w) 7e-05
 [FUNCAT] 30.02 organization of plasma membrane (S. cerevisiae, YHR005c) 1e-04
 [FUNCAT] 30.07 organization of endoplasmic reticulum (S. cerevisiae, YKL154w)
 1e-04
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 (S. cerevisiae, YHR005c) 1e-04
 [FUNCAT] 10.05.07 g-proteins (S. cerevisiae, YHR005c) 1e-04
 [FUNCAT] 06.04 protein targeting, sorting and translocation (S. cerevisiae, YKL154w)
 1e-04
 [FUNCAT] 08.19 cellular import (S. cerevisiae, YML001w) 3e-04
 [BLOCKS] BL00395A Alanine racemase pyridoxal-phosphate attachment site proteins
 [BLOCKS] BL01019B ADP-ribosylation factors family proteins
 [BLOCKS] BL01019A ADP-ribosylation factors family proteins
 [BLOCKS] BL01020D SAR1 family proteins
 [BLOCKS] BL01020C SAR1 family proteins
 [BLOCKS] BL01020B SAR1 family proteins
 [BLOCKS] BL01020A SAR1 family proteins
 [SCOP] dlplj_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 7e-36
 [SCOP] dlguua_ 3.25.1.3.10 Rap1A [Human (Homo sapiens) 8e-40
 [SCOP] dlrrf_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattus 2e-55
 [SCOP] dlhurb_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Homo 1e-58
 [SCOP] dlgota2 3.25.1.3.3 (1-54,171-326) Transducin (alpha subunit) [ra 2e-33
 [SCOP] dltdb2 3.25.1.3.2 (1-30,152-316) Transducin (alpha subunit 6e-36
 [PIRKW] glycoprotein 4e-19
 [PIRKW] monomer 1e-16
 [PIRKW] P-loop 3e-64
 [PIRKW] lipoprotein 4e-19
 [PIRKW] GTP binding 3e-64
 [SUPFAM] ADP-ribosylation factor 5e-22
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 3
 [PROSITE] SAR1 1
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D

SEQ MSFIFEWIYNGFSSVLQFLGLYKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPTLHPT
 1hurATTTTTCC

SEQ SEELTIAGMTFTTDFDLGGHEQARRVWKNYLPAINIGVFLVDCADHSRLVESKVELNALMT
 1hurA EEEEEETEEEEETTTTTTCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC

SEQ DETISNVPIILGNKIDRTDAISEEKLREIFGLYGQTTGKGNVTLKELNARPMVEVFMCSV
 1hurA TTTTTTTEEEEEETTTTTTCCCHHHHHHHHCGG.....

SEQ LKRQGYGEGFRWLSQYID
 1hurA

Prosite for DKFZphfkd2_46m4.3

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	60->64	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	111->115	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00017	32->40	ATP_GTP_A	PDOC00017
PS01020	171->197	SAR1	PDOC00782

Pfam for DKFZphfkd2_46m4.3

HMM_NAME ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)

HMM *GMgWfsIFrkMWGLWNKEMRILMLGLDNAGKTTILYMLKlgEIVTTIPT
 ++ FS+++++GL++K+++++LGLDNAGKTT+L+MLK++++ ++PT

Query 9 -YNGFSSVLQFLGLYKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPT 56

HMM IGFNVETVeYKNIKFNVVDVGGQdsIRPYWRHYYPNTDGIWVVDsADRD
 +++++E++++ ++F+++D+GG++++R++W++Y P+++GI+++VD+AD++

Query 57 LHPTSEELTIAGMTFTTFFDLGGHEQARRVWKNYLPAINGIVFLVDCADHS 106

HMM RMeEaKqELHaMLNEEELrDAPLlIFANKQDLpgAMsesEIREaLGLHeI
 R+ E+K+EL+A++++E ++++P+LI++NK+D+ +A+SE++RE+ GL+ +

Query 107 RLVESKVELNALMTDETISNVPIILGNKIDRTDAISEEKLREIFGLYGQ 156

HMM RCn.....RPWYIQMCCAvtGEGLYEGMDWLSNYInkRkK*
 +++ RP++++MC+++++G++EG++WLS+YI

Query 157 TTGKGNVTLKELNARPMVEFMCSVLKRQGYGEGFRWLSQYI----- 197

DKFZphfkd2_47a4

group: transcription factor

DKFZphfkd2_47a4.1 encodes a novel 280 amino acid protein with similarity to zinc finger proteins.

The new protein is a putative transcription factor with one C2H2 zinc fingers.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to C.elegans F46B6.7

potential frame shift at 1092, will be checked see BLASTX

Sequenced by MediGenomix

Locus: map="7q31"

Insert length: 1756 bp

Poly A stretch at pos. 1737, no polyadenylation signal found

```

1 CCCTTTTCTT TTCTGCCGGG TAATGGCTGC TTCCAAGACC CAGGGGGCTG
51 TCGCCCCGAAT GCAGGAAGAC CGTGATGGGA GCTGCAGCAC AGTCGGGGGT
101 GTAGGTTATG GGGTAAGGAT TGTATCCTGG AGCCGCTTTC CCTGCCAGAA
151 AGTCCAGGTG GCACCAACAC TTTAGAAGGT TCTCCATCTG TGCCTTGAT
201 TTTCTGTGAA GAACATTTTC CTGTGGCTGA ACAAGACAAA CTCTGAAGC
251 ACATGATTAT TGAGCATAAG ATTGTCATAG CTGATGTCAA GTTGGTTGCT
301 GATTTCCAAA GGTACATTTT ATATTGGAGG AAAAGGTTCA CTGAACAGCC
351 CATCACAGAT TTTTGTAGTG TAATAAGAA TAATTCCACT GCTCCATTTG
401 AAGAACAAAG GAATTATTTT TTGTTATGTG ACGTTTTACC AGAAGATAGA
451 ATTCTTAGAG AAGAGCTTCA GAAACAGAGA CTGAGAGAAA TTCTGGAAAC
501 ACAGCAGCAA GAACGAAATG ATAACAATTT TCATGGCGTT TGTATGTTTT
551 GCAATGAAGA ATTCCTTGGA AACAGATCTG TTATTTTGAA CCACATGGCC
601 AGAGAACATG CTTTCAACAT TGGATTGCCA GACAACATTG TAAACTGCAC
651 TGAATTTTGT TGTACATTAC AGAAAAAGCT TGACAATTG CAGTGCCTGT
701 ACTGTGAGAA GACCTTCAGG GGCAAAAATA CACTTAAAGA TCACATGAGG
751 AAAAAACAGC ATCGTAAGAT TAATCCTAAG AACAGAGAAT ATGACAGATT
801 TTATGTCATC AATTATTTGG AACTTGGAAG ATCGTGGGAG GAAGTTCAGT
851 TGGAAGATGA TCGGGAGTTG CTGGACCATC AGGAAGATGA CTGGTCTGAT
901 TGGGAAGAAG ACCCTGCCTC TGCAGTCTGC TTATTTTGTG AAAAGCAAGC
951 AGAACAATT GAGAAAGTTG ATGTCCACAT GGAGGATGCA CACGAATTTG
1001 ATCTTCTCAA AATAAAGTCA GAACTTGGAT TAAATTCTA TCAGCAAGTG
1051 AAACCTGGTA ATTTTATTCG GAGGCAAGTT CACCAATGCA GATGATGGCT
1101 GCCATGTGAA GTTCAAATCC AAAGCAGACT TAAGAACTCA CATGGAAGAA
1151 ACTAAACACA CTTGCTGCTC CCCCAGATGA AAGACGTGGG ATCAACTGGA
1201 GTATTATTTT CCAACCTATG AAAATGACAC TCTCCTGTGT AACTATCTG
1251 ACAGTGAAAG TGACCTGACA GCTCAGGAAC AAAATGAAAA TGTTCCTATC
1301 ATCAGTGAAG ATACATCTAA ACTGTATGCT TTGAAACAAA GCAGTATTTT
1351 GAACCAAGTG CTAATAAAG AGTACTTGAA AACCTAGAAG AACTACCAC
1401 AGAAGCAATT TTTATGTTT TTCTCCTATG AGACAGATAT GAAAGAACAA
1451 TTTAAATTTG AACATCAACA AAAGATTGGT CCTTGGTGAA ATAAACTTTT
1501 CAAAAATGAA TGTTCTTTTC AAAAAATAAA GTAGAAAAAT GCACTTACTA
1551 AGAACATGAA AAAAAATGA AGTAGGAAAA TAAGATGAAG ACTTTGTATT
1601 TTGGCTGTAA AGTTTATTG TGTGATCATC TTAAATTATC TCACTTCATT
1651 AAACATATAA TTATATATAG AAGTATATGT CAATTACAAA GAAATGAAAT
1701 GTTCAAATTA TTTATAAACC TGATTTTCA ATCAGCGAAA AAAAAAATAA
1751 AAAAAA

```

BLAST Results

Entry AC004112 from database EMBL:
Homo sapiens BAC clone RG313E03 from 7q31, complete sequence.
Score = 2660, P = 3.0e-241, identities = 534/535
> 10 exons

Entry AC004111 from database EMBL:
Homo sapiens BAC clone RG103H13 from 7q31, complete sequence.
Score = 598, P = 5.8e-17, identities = 128/137
1 exon

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 253 bp to 1092 bp; peptide length: 280
Category: similarity to unknown protein

```

1 MIIHKIVIA DVKLVADFQR YILYWRKRFT EQPITDFCSV IRINSTAPFE
51 EQENYFLLCD VLPEDRILRE ELQKQRLREI LEQQQQERND NNFHGVCMFC
101 NEEFLGNRSV ILNHMAREHA FNIGLPDNIV NCNEFLCTLQ KKLDNLQCLY
151 CEKTFRGKNT LKDHMRKKQH RKINPKNREY DRFYVINYLE LGKSWEDEVQL
201 EDDRELLDHQ EDDWSDWEH PASAVCLFCE KQAEITIEKLY VHMDAHEFD
251 LLKIKSELGL NFYQQVKLVN FIRRVHQCR

```

BLASTP hits

Entry CEF46B6_6 from database TREMBLNEW:
product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
>TREMBL:CEF46B6_6 product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
Score = 630, P = 1.1e-61, identities = 123/289, positives = 183/289

Entry AF059531_1 from database TREMBLNEW:
gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds. >TREMBL:AF059531_1 gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds.
Score = 120, P = 1.5e-04, identities = 23/78, positives = 42/78

Entry YB9M_YEAST from database SWISSPROT:
34.7 KD PROTEIN IN SHM1-MRPL37 INTERGENIC REGION.
Score = 112, P = 4.6e-04, identities = 43/165, positives = 71/165

Alert BLASTP hits for DKFZphfkd2_47a4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_47a4, frame 1

Report for DKFZphfkd2_47a4.1

```

[LENGTH]      280
[MW]           33921.94
[pI]           5.63
[HOMOL]        TREMBL:CEF46B6_5 gene: "F46B6.7"; Caenorhabditis elegans cosmid F46B6 1e-56

[BLOCKS]       BL01032B Protein phosphatase 2C proteins
[BLOCKS]       BL00028 Zinc finger, C2H2 type, domain proteins
[PROSITE]      MYRISTYL 1
[PROSITE]      ZINC_FINGER_C2H2 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 3
[PROSITE]      TYR_PHOSPHO_SITE 2
[PROSITE]      PKC_PHOSPHO_SITE 2
[PROSITE]      ASN_GLYCOSYLATION 2
[PFAM]         Zinc finger, C2H2 type
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY 8.21 %

```

```

SEQ  MIIHKIVIA DVKLVADFQR YILYWRKRFT EQPITDFCSV IRINSTAPFE EQENYFLLCD
SEG  .....
PRD  cccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeeccccccchhhhhheeeccc

SEQ  VLPEDRILRE ELQKQRLREI LEQQQQERND NNFHGVCMFC NEEFLGNRSV ILNHMAREHA
SEG  ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeeccccccceehhhhhhhh

SEQ  FNIGLPDNIV NCNEFLCTLQ KKLDNLQCLY CEKTFRGKNT LKDHMRKKQH RKINPKNREY

```



```

SEG .....
PRD hccccccccchhhhhhhhhhhhhheeeccccccchhhhhhhhhcccccccc

SEQ DRFYVINYLELGKSWEEVQLEDDRELLDHQEDDWSDEEHPASAVCLFCEKQAETIEKLY
SEG .....
PRD ceeeeeeeccccchhhhhhhcchhhhhccccccccccccccccchhhhhhhhhhhhh

SEG VHMEDAHEFDLLKIKSELGLNFIYQQVKLVNFIRQVHOCR
SEG .....
PRD hhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhcccc

```

Prosites for DKF2phfkd2_47a4.1

PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	107->111	ASN_GLYCOSYLATION	PDOC00001
PS00004	27->31	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	160->164	CK2_PHOSPHO_SITE	PDOC00006
PS00006	194->198	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00007	178->185	TYR_PHOSPHO_SITE	PDOC00007
PS00007	13->22	TYR_PHOSPHO_SITE	PDOC00007
PS00008	124->130	MYRISTYL	PDOC00008
PS00028	148->171	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKF2phfkd2_47a4.1

HMM_NAME	Zinc finger, C2H2 type		
HMM	*CpwpDCgKtFrrwsNLrRHMR..T.H*		
	C +	C+KTFR + +L+ HMR	H
Query	148	CLY--CEKTFRGKNTLKDHMRKK-QH	170

DKFZphfkd2_4b6

group: kidney derived

DKFZphfkd2_4b6 encodes a novel 133 amino acid protein with similarity to Homo sapiens clone 25003 partial CDS.

No informative BLAST results: No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to Homo sapiens clone 25003

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1936 bp

Poly A stretch at pos. 1916, polyadenylation signal at pos. 1890

```
1 GGGAGACTTG CAATGAAGTT AGAATGAACA GGAGGAGTCT GCAGCTTTTC
51 AGTGCCTGGG ATAACATATAG TTAAAGATC ATTGTGTAAA ATAGGATTTT
101 TAGTCAGCAT GCATTGTTTT AAACCGACTA ACTGATAGCC TAAACTTTTA
151 TTTTTCGATT TTGCCAATCC TTGGAGTTTT GTTTTGCAGA ATTAAGAAAA
201 AAATGAATGT ATGATCATCT GAAAAGGGCT TTCTCTCAAT CCCACTTCAT
251 GGCATGACCT CTGCTGGATC ATTAGTTCTA GCCAGAGAAG TAGCAAAGGA
301 ACATGACGTC TGAGACCTCC CTTCCCTCAT CAGTGGGGCT GACTGAGCTG
351 GGGGCTTGAA GCCGGAGGTA ACCTTTCCTG TCGAATGTTT CTTTAGAGAA
401 TGGCAATGGT CTCTGCGATG TCCTGGGTCC TGTATTTGTG GATAAGTGCT
451 TGTGCAATGC TACTCTGCCA TGGATCCCTT CAGCACACTT TCCAGCAGCA
501 TCACCTGCAC AGACCAGAAG GAGGGACGTG TGAAGTGATA GCAGCACACC
551 GATGTTGCAA CAAGAATCGC ATTGAGGAGC GGTCACAAAAC AGTAAAGTGT
601 TCCTGTCTAC CTGGAAAAGT GGCTGGAACA ACAAGAAACC GGCCTTCTTG
651 CGTCGATGCC TCCATAGTGA TTGGAAAATG GTGGTGTGAG ATGGAGCCTT
701 GCCTAGAAGG AGAAGAATGT AAGACACTCC CTGACAATTC TGGATGGATG
751 TGCGCAACAG GCAACAAAAT TAAGACCACG AGAATTCACC CAAGAACCTA
801 ACAGAAGCAT TTGTGGTAGT AAAGGAAAAC CAACCTCTG GAAAATACAT
851 TTTGAGAATC TCAACATCTC CACATATATA CAAGCCAAAT GGATTTCTTA
901 CTTGCACTTT GACTGGCTAC CAGATAATCA CAGTGCCTTT AGTGTGTGTA
951 ACGAAATATC CTACAGTGAG AAGACACAGC GTTTTGGCAT CACCATGGAA
1001 AGTGGGCTTA AAAAAGGGTC TTCTCAGTGA AATTTTGGG CATCATGAAG
1051 AACGATCAAC TATCTTCTAA TTTGAATCTA TAGTTACTTT GTACCATTTG
1101 AAATATATGT ATATATATAT ATATAATATT TTGAAATATT ATCTATTCTC
1151 TTCAAGAAAT GAACAGTACC ACAGTTTGAG ACGGCTGGTG TACCCCTTTG
1201 AGTTTTGGAT GTTTTGTCTG TTTTGTCTTG TTTTGTAGT CATTTCTTTT
1251 TCTAACGGCA AGGAAGATAT GTGCCCTTTT GAGAATTCAG GATGGCACTG
1301 ACACGGGAAG GCCAGCTACA GGTGGACTCC TGGAAATTGA GGCATCATAA
1351 TGATACTGAA TCAAGAACTT CCTTCTGCTT CTACCAGATG GCCCAAGGAA
1401 GCACATCGTC CTGTTTTTAT GCTTCTTACC CTGTGCAATA TTAGCATGCA
1451 ACCTTGGCTT ACATAGTCAT ACTTTATATT CAATTGATAT ATAATAACCG
1501 TTCTAACCTC TTCCAGGAAA ATATTTTTAG AACTACTAGC TTTTCCACTT
1551 AGAAGAAAAT GAGGATTCTT AAGGGAGCCA CTCCACCATG CTATTAAGAC
1601 TCTGGCAGAG TTATGGGTAG GATATGGATC CCTACATGAA TAAGTCCTGT
1651 AAATCAATG TCTTAAGGCT TTGTATAGCT GTCCTAGACT GCAGAAATGT
1701 CCTCTGATTA AATCCAAAGT CTGGCATCGT TAACTACATA GTGCTGTAGC
1751 AACCAAGTCT ATCATGGCAT CTCTTCTAT GTTTGGTTTG CTTTTTCCAA
1801 GAGTATTCAG GTCTCTCTT GTGAGATAGG AAGGCCATGA AAACAATTAG
1851 ATTTCAAGAT GATCTATGTG ACCAAATGTT GGACAGCCCT ATTAAGTGG
1901 TAAACAACCT CTTTCTAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for Name:

1 MAMVSAMSWV LYLWISACAM LLCHGSLQHT FQQHHLRPE GGTCEVIAAH
51 RCCNKNRIEE RSQTVKCSCL PGKVAGTTRN RPSCVDASIV IWKWWCEMEP
101 CLEGECKTL PDNSGWMCAT GNKIKTTRIH PRT

No BLASTP hits available

TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA sequence, partial cds., N = 1, Score = 242, P = 1.7e-20

HSPs :

Query: 42 GTCEVIAAHRCCNKNRIEERSQTVKSCSLPGKVAGTTRNRPSVDASIVWKWCCMEPC 101
GTCE++ R ++ R QT +C+C G++AGTTR RP+CVDA I+ K WC+M PC
Sbjct: 76 GTCEIVTLDRDSSQPRRTIARQTARCACRKGQIAGTTRARPACVDARIKTKQWCDMLPC 135

```

Query:   102 LEGEECKTLPDNSGWMCAT-GNKIKTTRI 129
          LEGE C  L + SGW C  G +IKTT +
Sbjct:   136 LEGEGCDLLINRSGWTCTQPGGRIKTTTV 164

```

```
[LENGTH]      133
[MW]           15030.64
[PI]           8.49
[HOMOL]        TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
sequence, partial cds. 4e-20
[KW]           Alpha_Beta
[KW]           SIGNAL PEPTIDE 26
```

```

SEQ      MAMVSAMSWLYLWISACAMLLCHGSLQHTFQOHHLHRPEGGTCEVIAAHRCNKNRIEE
PRD      cccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhccccccccceeeeeccccchhh

SEQ      RSQTVKCSCLPGKVAGTTRNRPSCVDASIVIKWKCMEPECLGECKTLPDNSGWMCAT
PRD      hhhhhhhccccccccccccccccccccceeeehhhhhhhccccccccceeeccccceec

SEQ      GNKIKTTRIHPRT
PRD      cccccccccccccc

```

(No Prosite data available for DKFZphfkd2_4b6.1)

(No Pfam data available for DKFZphfkd2_4b6.1)

DKFZphfkd2_4c8

group: kidney derived

DKFZphfkd2_4c8 encodes a novel 153 amino acid protein with partial similarity to huntington's associated protein HAP1.

The novel protein contains a leucine zipper involved in protein-protein interaction.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KIAA0549 and HAP1

potential frame shift at Bp -1350-1500 will be checked

Sequenced by GBF

Locus: unknown

Insert length: 3182 bp

Poly A stretch at pos. 3162, polyadenylation signal at pos. 3135

```
1 GGGCTTCCCC CATAGAATTT TTCTTTTCAT TGCCCACTTT ACTGTTTTGG
51 CTCCAGACTG TCGTTAAGAA TGTACAGCCT AATTCTGGTG TGTTCGGGA
101 TATTCTTCTG TCCAGTATTC TGAAGGGCG GGGAGGCATG GCAGCGTTTT
151 ACTTGACGTT GATGGTGCTG TGAAGTCCAT TCTTCTCTCT SCAAGACTAC
201 TGAATATGCA GAAATTTATC GAAGCGGATT ATTATGAAC AGACTGGTAT
251 TATGAAGAAT GCTCGGATGT TTTATGTGCT GAAAGAGTTG GCCAGATGAC
301 TAAGACATAT AATGACATAG ATGCTGTAC TCGGCTTCTT GAGGAGAAAG
351 AGCGGGATTT AGAATTGGCC GTCGCATCG GCCAGTCGTT GTTGAAGAAG
401 AACCAAGACCC TAACCGAGAG GAACGAGCTG CTGGAGGAGC AGGTGGAACA
451 CATCAGGGAG GAGGTGTCTC AGCTCCGGCA TGAGCTGTCC ATGAAGGATG
501 AGCTGCTTCA GTTCTACACC AGCGCAGCGG AGGAGAGTGA GCCCGAGTCC
551 GTTTGGTCAA CCCCGTTGAA GAGGAATGAG TCGTCCTCCT CAGTCCAGAA
601 TTACTTTTCAT TTGGATTCTC TTCAAAAGAA GCTGAAAGAC CTTGAAGAGG
651 AGAATGTTGT ACTTCGATCC GAGGCCAGCC AGCTGAAGAC AGAGACCATC
701 ACCTATGAGG AGAAGGAGCA GCAGCTGGTC AATGACTGCG TGAAGGAGCT
751 GAGGGATGCC AATGTCCAGA TTGCTAGTAT CTCAGAGGAA CTGGCCAAGA
801 AGACGGAAGA TGTCGCCGCG CAGCAAGAGG AGATCACACA CCTGCTATCG
851 CAAATAGTTG ATTTGCAGAA AAAGGCAAAA GCTTGCCGAG TGGAAAATGA
901 AGAAGCTTGT CAGCATCTGG GGGCTGCTAA GGATGCCAG CGGCAGCTCA
951 CAGCCGAGCT GCCTGAGCTG GAGGACAAGT ACGCAGAGTG CATGGAGATG
1001 CTGCATGAGG CGCAGGAGGA GCTGAAGAAC CTCGGAACA AAACCATGCC
1051 CAATACACAG TCTCGGCGCT ACCACTCACT GGGCTGTTT CCCATGGATT
1101 CCTTGGCAGC AGAGATTGAG GGAACGATGC GCAAGGAGCT GCAGTTGGAA
1151 GAGGCCGAGT CTCCAGACAT CACTCACCAG AAGCGTGTCT TTGAGACAGT
1201 AAGAAACATC AACCAGGTTG TCAAGCAGAG ATCTCTGACC CCTTCTCCCA
1251 TGAACATCCC CGGCTCCAAC CAGTCCTCGG CCATGAACTC CCTCCTGTCC
1301 AGCTGCGTCA GCACCCGCCG GTCCAGCTTC TACGGCAGCG ACATAGGCAA
1351 CGTCGTCTCT GACAACAAGA CCAACAGCAT CATCTGGAA ACAGAGGCAG
1401 CCGACCTGGG AAACGATGAG CGGAGTAAGA AGCCGGGGAC GCCGGGCACC
1451 CCCAGGCTCC CACGACCTGG AGACGGCGCT GAGCGGGCTG TCCCTGCGCC
1501 GGGAGAACTA CCTCTCGGAG AGGAGGTTCT TTGAGGAGGA GCAAGAGAGG
1551 AAGCTCCAGG AGCTGGCGGA GAAGGGCGAG CTGCGCAGCG GCTCCCTCAC
1601 ACCCACTGAG AGCATCATGT CCCTGGGCAC GCACTCCCGC TTCTCCGAGT
1651 TCACCGGCTT CTCTGGCATG TCCTTCAGCA GCCGCTCCTA CCTGCCTGAG
1701 AAGCTCCAGA TCGTGAAGCC GCTGGAAGGT GATCACGCGG GGCCTCGGCC
1751 CCTCTCTGTC CTCTGGGGG ACTCCCTTTG GTCCCTGATC CACCTGCGGA
1801 AGGCGGGGCA CCTCTGTAC GCCTACTCCT TTTTCTTCCG CGACAGCCAC
1851 CCGCGCTGCT GGTTTGAGTT CCTCTGAGGG TGGTGCTCAG CCTAGGCCTC
1901 CGTCCCTCCC CTCTGGCTGG CAGGTGTGAC AATGCACACA TAGGCCATGA
1951 AACTCGCCGA GGAAAGACAA GCATGTGCAC TGTGGTCTTC TAGTCTTTT
2001 CTTTGCTTTT AGAACCTTAG AAATAAAAC TTTTGTGGCG GTAGAGGCAC
2051 TGCTAACTGA TTCAAAAATT AATTAGGTTT TGCTGTGGG TGTGAGGAAT
2101 GCAGAAAAAT AATGCTTTAG CTTTCTGCA GTTTTGGTGT CGGGGAGAGG
2151 TTCCAAGCAA ACTCTATTAA ATGGGGATT TTTTTCCTCC ATAACCACCT
2201 GAATGTGATT TGTGGGCTTA TGTGTTCTGA TTTGAATTC ATATAGCAAG
2251 GTTGTGGCTT TTGGCAGATG CAGTATGTTG TGAGCGCGGC TCCTAGAGTC
2301 TACAATTTGG AGTCCAGGAA GGGGTGGCTG TGGAGACAAG TGAGTTTTGT
2351 ACCTCCGTAA GCCACCCTTT TTCAGGGTCA GTTCAATGTT TAGTATCAGG
2401 GGCATCTCAG ATGATTAAAC TCATGGGAAA AACTTCCTCC TTCCTCTCT
2451 CCCTCTTGCC CTCTGCTCTC TTTTCTTTT TTTTCTTTT AATTGTTGGC
2501 CTTATAAAAT GTTTTCCCTC TACCTGCTGC TACTCTGCCA AGAGCCACCA
2551 AGTGCTTATA TTTTTCATTT TTTACTCCT TAGTTTGGAA AGCCATATAC
2601 GTTTGAGAAG GTGTTTAAA ACTCTGTGTT ACACCTACGA TGCAAAGCCA
2651 AATCAGAACT TCTGTAAGGC AGAAGCTTCC CAACCTTAAA AAAATTATTG
```



```

2701 TCCCCTCTAG GAGCCTTCTT AGACGTTTTT TCCTAATCAC CCCCCAAGA
2751 CATTTTAATA CCACATATAT ATTGTTTATG TACTATATGT ATATACATAA
2801 ACAATACATA AGCAATACAT CTGTGGTATT AAAATTAAAA AGAATCCAAT
2851 TATGTTTACC TCAAAAGAAC CTGTTTTTGC TTCTTGGGAG CAATATTGCC
2901 CCTGTGAGAC TGCATGCTAT AAGGTAAGGT TGTGCTTGTT AAAGACCCAA
2951 GACATGACTG GGTTCACAG TCTCCAAAGG AAGAGGGTGG GCTAGTTTGT
3001 TTTTATTATT ATTTTAAAT TGTATAATTG GGGTCTTTCT TAGAGTTCAG
3051 AAAAGGTATA GCTTACTCTT TTTTAATTGT TTATTAGTT GTAAGCTTAG
3101 TGATTGTTTT CTGATCCACA TTGTGTGTGT TCTCAATAA AATCTTCAT
3151 TTCTGCAATT TAAAAA AAAA AAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 206 bp to 1531 bp; peptide length: 442
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (139-161)

```

1 MQKFIEADYY ELDWYEECS DVLCAERVGO MTKTYNDIDA VTRLLEEKER
51 DLELAARIGQ SLLKKNKTLT ERNELLEEQV EHIREEVSQ L RHELSMKDEL
101 LQFYTSAAEE SEPESVCSTP LKRNESSSSV QNYFHLDLQ KKLKDLLEEN
151 VVLRSEASQL KTETITYEEK EQQLVNDVCK ELRDANVQIA SISEELAKKT
201 EDAARQEEI THLLSQIVDL QKAKACAVE NEELVQHLGA AKDAQRLTA
251 ELRELEDKYA ECMEMLEHAQ EELKNLRNKT MPNTTSRRYH SLGLFPMDSL
301 AAEIEGTMRK ELQLEEAESP DITHQKRVFE TVRNINQVVK QRSITPSPMN
351 IPGSNQSSAM NSLLSSCVST PRSSFYGS DI GNVVLDNKTN SIILETEAAD
401 LGNDERSKKP GTPGTPRLPR PGDGAEAAVP APGELPLGEE VL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphkd2_4c8, frame 2

PIR:S72555 huntingtin-associated protein HAP1 - human (fragment), N = 1, Score = 234, P = 8.6e-19

TREMBL:CEUT27A3 7 gene: "T27A3.1"; Caenorhabditis elegans cosmid T27A3., N = 1, Score = 226, P = 9.9e-16

PIR:S67495 huntingtin-associated protein HAP1-A - rat, N = 1, Score = 215, P = 1.6e-14

>PIR:S72555 huntingtin-associated protein HAP1 - human (fragment)
 Length = 320

HSPs:

Score = 234 (35.1 bits), Expect = 8.6e-19, P = 8.6e-19
 Identities = 66/189 (34%), Positives = 110/189 (58%)

```

Query: 109 EESEPEVCSTPLKRNE--SSSSVQNYFH---LDSLQKKLKDLEENVVLRSEASQLKTE 163
      EE+E + C+ P + S ++ + H L++LQ+KL+ LEEEN LR EASQL T
Sbjct: 28 EEAEEDLQCAHPCDAPKLISQEALLHQHHCPLQLEALQEKLRLEENHQLREEASQLDT- 86

Query: 164 TITYEEKEQQLVNDVCKELRDANVQIASISEELAKKTEDAARQEEIETHLLSQIVDLQKK 223
      E++EQ L+ +CV++ +A+ Q+A +SE L + E+ RQQ+E+ L +Q++ LQ++
Sbjct: 87 ---LEDEEQMLILECVEQFSEASQMAELSEVLVLRLENYERQQQEVARLQAVLKLQQR 143

Query: 224 AKACAVENEELVQHLGAQKDAQRLTAEL--LRELEDKYAECE--MLHEAQEELKNL-RN 278
      + E E+L + L + K+ Q QL E L ++ AE + + + + + RN

```


Sbjct: 144 CRMYGAETEKLOKQLASEKEIQMQLQEEETLPGFQETLAEELRTSLRRMISDPVYFMERN 203
 Query: 279 KTMP--NTTSRRY 289
 MP +T+S RY
 Sbjct: 204 YEMPRGDTSSLRY 216

Peptide information for frame 3

ORF from 1416 bp to 1874 bp; peptide length: 153
 Category: similarity to known protein
 Classification: unset

1 MSGVRSRGRR APPGSHDLET ALRRLSLRRE NYLSERRFFE EEQERKLQEL
 51 AEKGLRSGS LPTESIMSL GTHSRFSEFT GFSGMSFSSR SYLPEKLQIV
 101 KPLEGDHAGP RPLSVLLGDS LWSLIHLRKA GHLCHAYSFF FRDSHPRCWF
 151 EFL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4c8, frame 3

TREMBL:AB011121.1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds., N = 1, Score = 252, P = 5.5e-21

>TREMBL:AB011121.1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds.
 Length = 469

HSPs:

Score = 252 (37.8 bits), Expect = 5.5e-21, P = 5.5e-21
 Identities = 57/98 (58%), Positives = 69/98 (70%)

Query: 8 GRRAPPGSHDLETALRRLSLRRENYLSERRFFEEEQERKLQELAEGELRSGSLTPTESI 67
 G+ P G DL TAL RLSLRR+NYLSE++FF EE +RK+Q LA++ E SG +TPTES+
 Sbjct: 27 GQPGPSGSDSLATALHRLSLRRQNYLSEKQFFAEWQRKIQLVADQKEGVSGCVTPTESL 86
 Query: 68 MSLGTHSRFSEFTGFSGMSFSSRSYLPEKLQIVKPLEG 105
 SL T SE T S S R ++PEKLQIVKPLEG
 Sbjct: 87 ASLCTTQ--SEITDLSSAS-CLRGFMPEKLQIVKPLEG 121

Pedant information for DKFZphfkd2_4c8, frame 2

Report for DKFZphfkd2_4c8.2

[LENGTH] 442
 [MW] 50020.14
 [pI] 4.77
 [HOMOL] TREMBL:AF040723_1 product: "neuroan1"; Homo sapiens neuroan1 mRNA, complete
 cds. 5e-29
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
 5e-08
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YIL149c] 5e-08
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-08
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YIL138c]
 6e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR130c] 2e-07
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-06
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-06
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 jannaschii, MJ1643] 1e-06
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 -
 myosin-1 isoform] 3e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YKR095w] 4e-06
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 4e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 2e-05
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 2e-05


```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c]
5e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
1e-04
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 1e-04
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YNL272c] 3e-04
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YNL272c] 3e-04
[BLOCKS] BL01289B
[BLOCKS] BL00415M Synapsins proteins
[EC] 3.6.1.32 Myosin ATPase 2e-07
[PIRKW] tandem repeat 2e-07
[PIRKW] heterodimer 1e-06
[PIRKW] endocytosis 9e-07
[PIRKW] heart 1e-06
[PIRKW] transmembrane protein 4e-07
[PIRKW] zinc finger 9e-07
[PIRKW] metal binding 9e-07
[PIRKW] DNA binding 3e-06
[PIRKW] muscle contraction 2e-07
[PIRKW] acetylated amino end 3e-06
[PIRKW] actin binding 2e-07
[PIRKW] mitosis 1e-06
[PIRKW] microtubule binding 1e-06
[PIRKW] ATP 2e-07
[PIRKW] chromosomal protein 1e-06
[PIRKW] receptor 3e-08
[PIRKW] thick filament 2e-07
[PIRKW] phosphoprotein 8e-06
[PIRKW] glycoprotein 3e-08
[PIRKW] skeletal muscle 3e-06
[PIRKW] DNA condensation 1e-06
[PIRKW] alternative splicing 2e-06
[PIRKW] coiled coil 2e-07
[PIRKW] P-loop 2e-07
[PIRKW] heptad repeat 4e-07
[PIRKW] methylated amino acid 2e-07
[PIRKW] peripheral membrane protein 9e-07
[PIRKW] cardiac muscle 6e-06
[PIRKW] hydrolase 2e-07
[PIRKW] muscle 2e-06
[PIRKW] cytoskeleton 2e-06
[PIRKW] Golgi apparatus 4e-07
[PIRKW] calmodulin binding 9e-07
[SUPFAM] myosin motor domain homology 2e-07
[SUPFAM] tropomyosin TPM1 2e-06
[SUPFAM] giantin 4e-07
[SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
[SUPFAM] human early endosome antigen 1 9e-07
[SUPFAM] unassigned kinesin-related proteins 4e-07
[SUPFAM] M5 protein 8e-08
[SUPFAM] cytoskeletal keratin 3e-06
[SUPFAM] myosin heavy chain 2e-07
[SUPFAM] conserved hypothetical P115 protein 1e-06
[SUPFAM] centromere protein E 1e-06
[SUPFAM] pleckstrin repeat homology 2e-06
[SUPFAM] kinesin motor domain homology 4e-07
[PROSITE] LEUCINE_ZIPPER 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 6.79 %
[KW] COILED_COIL 27.15 %

```

```

SEQ      MQKFIEADYYELDWYEECSVDLCAERVGOMTKTYNDIDAVTRLLEEKERDLELAARIGQ
SEG                                             .XXXXXXXXXXXXXXXXXX...
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....C

SEQ      SLKKKNKTLTERNELLEQVEHIREEVSQLRHESMKDELLQFYTSAAEESPEPVCSTP
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      LKRNESSSVQNYFHLDSLQKKLKOLEENVLRSEASQLKTETITITYEEKQQLVNDCVK
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
```


Prosites for DKFZphfkd2_4c8.2

PS00029 139->161 LEUCINE ZIPPER PDOC00029

(No Pfam data available for DKFZphfkd2_4c8.2)

Pedant information for DKFZphfkd2_4c8, frame 3

Report for DKFZphfkd2_4c8.3

```
[LENGTH]      153
[MW]           17642.03
[pI]           9.38
[HOMOL]        TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens
mRNA for KIAA0549 protein, partial cds. 2e-12
[KW]           Alpha_Beta
[KW]           LOW COMPLEXITY      12.42 %
```

```
SEQ      MSGVRSRGRRAPPGSHDLETALRLRLSLRRENYLSERRFFEEEQERKLQELAEKGLRSGS
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
```



```
SEQ      LTPTESIMSLGTHSRFSEFTGFSGMSFSRSSYLPPEKLQIVKPLEGDHAGRPPLSVLLGDS
SEG      .....
PRD      ccccceeeccccceeeccccccccccccccccchhhhhhhhccccccccccccceeeccc
```



```
SEQ      LWSLIHLRKAGHLCHAYSFFFRDSDSPRCWFEFL
SEG      .....
PRD      chhhhhhhhhccccceeeeeecccccccccccc
```

(No Prosite data available for DKFZphfkd2_4c8.3)

(No Pfam data available for DKFZphfkd2_4c8.3)

DKFZphfkd2_4k14

group: intracellular transport and trafficking

DKFZphfkd2_4k14.3 encodes a novel 254 amino acid putative GTP-binding protein nearly identical to Rab6.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory (biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes.

rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport.

The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

strong similarity to Rab6

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3084 bp

Poly A stretch at pos. 3061, polyadenylation signal at pos. 3043

```
1 GGGGCACTCA GCAGGTTGGG CTGCGGCGGC GCGGGCTGGG GAAGCCGAAG
51 CGCCGCGCGT GAGAGATCCC GGATACATCT GCGGTTTGGG CTCCGCCACC
101 CTCCGCTCTCT CTCCCGCAGG TCTCTGAGCC GGGTGCGGAA GGAGGGAACG
151 GCCCTAGCCT TGGGAAGCCA AAGCACACCC CTGGCTCCCG CCGACACCGC
201 CCTCCTTCCC TTCCAGCCG CGGGCCTCGC TCCGTGCTCG GCTACTCTGC
251 CGGGAGGCGG CGGCGGCTGC CAGTCTGTGG CGAGCCCTGC TGGCCTCCAG
301 CCGGGCTTCT CCAGCCGGGC TCCTCCACCG GCCCTTGCGG GGGCACAGAG
351 AGCTCGGCGC CGGCCCTTCC GCTCGCCTTT TTCGTGAGCC GGTGGAGGA
401 GCATCGGTCC GGGAGGTCTC TGGGCTGAGG CGGCGACAGC TCCTCTAGTT
451 CCACCATGTC CGCGGGCGGA GACTTCGGGA ATCCGCTGAG GAAATTCAAG
501 CTGGTGTTCC TGGGGGAGCA AAGCGTTGCA AAGACATCTT TGATCACCAG
551 ATTCAGGTAT GACAGTTTGG ACAACACCTA TCAGGCAATA ATTGGCATTG
601 ACTTTTATC AAAAATATG TACTTGGAGG ATGGAACAAT CGGGCTTCGG
651 CTGTGGGATA CGGCGGGTCA GGAACGTCTC CGTAGCCTCA TTCCAGGTA
701 CATCCGTGAT TCTGCTGCAG CTGTAGTAGT TTACGATATC ACAAAATGTA
751 ACTCATTCCA GCAAATACA AAGTGGATTG ATGATGTCAG AACAGAAAGA
801 GGAAGTGATG TTATCATCAC GCTAGTAGGA AATAGAACAG ATCTTGCTGA
851 CAAGAGGCAA GTGTCAAGTG AGGAGGGAGA GAGGAAAGCC AAAGGGCTGA
901 ATGTTACGTT TATTGAAACT AGGGCAAAAA CTGGATACAA TGTAAAGCAG
951 CTCTTTCGAC GTGTAGCAGC AGCTTTGCCG GGAATGGAAA GCACACAGGA
1001 CGGAAGCAGA GAAGACATGA GTGACATAAA ACTGGAAGAG CCTCAGGAGC
1051 AAACAGTCAG CGAAGGGGGT TGTTCCTGCT ACTCTCCCAT GTCATCTTCA
1101 ACCCTTCCTC AGAAGCCCCC TTACTCTTTC ATTGACTGCA GTGTGAATAT
1151 TGGCTTGAAC CTTTCCCTT CATTAATAAC GTTTTGCAAT TCATCATTGC
1201 TGCCGTGCTC GTGGAGGTGA TCTATTAGCT TCACAAGCAC AAAAAAGTC
1251 AGCGTGCTTA TTATTTATAT TTTACAAAAA GCCAAATTAT TTCAGCATAT
1301 TCCGGTGATA ACTTTAAAAA TTAGATACAT TTTCTTAACA TTTTCTTCTT
1351 TTTTAATGTT ATGATAATGT ACTTCAAAAT GATGGAATC TCAACAGTAT
1401 GAGTATGGCT TGGTTAACGA GCAGTATGTT CACAGCCTGC TTTATCTCTC
1451 CTTGCTCTTC TCACCTCTCC CTTACCCCGT TCCCTATTTC CGTGTCTTFA
1501 CTTAGCCTCC CCCCACCTTC TCAAAACAAA CAAGAGATGG CAAAGCAGCA
1551 GTCCGACCAA GCCCACTGGA ATTATCCTTT AATTTTACAG ATACCACTTG
1601 CTGTAGGCTG TGGACCAAGA TGTCCAGAAT TATTCTTGAG CACTGATGTA
1651 AATTACTTAG ATCTTCTTTG AGGTCAGAAT TCAGCGATCA CGGTAGGCAG
1701 TGCTTGAATG AGAAAAGCCT CCTGGTGCAT CTTCAAAATG AGTCCTAAAG
1751 AACATACTGA GTACTTATAA GTAGCAGAAC ATAAAATGTA TTTCTGACTA
1801 ACACAAATGG TCCTTTCACA TGTGCTTTAT TAGACTCTGG GAGAGAAAAG
1851 TAACCAAGTG CTTCAGAAC A GGTTTTGTAGT ATTTACTTCT TCATGGTAAG
1901 ATAATGAAGT TCTAATGAAC TATTTCTCCC AAGGTTTTAA AATTGTCAAG
1951 AGTTATTCTG TTTGTTTAAA AAGTAAGAAA CCTCTGTAAG CAATAGATTT
2001 TGCTTGGGTT TTCTTTCTTA AAAAAATAAT ACTATGCAGG CAAGACACCA
2051 TAAAGTTTAA ATTCTTTACA GAAGAACCAG TGGAAGAATT TAAATTTGGC
2101 ACTACGATCA AAACACTGTA ATTAGCAGAA ATAACGATAT CTAAGGCTTA
2151 CCAGCAAAAG AACCTCAGC AGAATAGCAA AAACCTTGCT CAGGACATTT
2201 GAGGTCAAAT TGAAGACGGA AGACGGAAAC CGGAACCCGT TTTCTGTAA
2251 GCCCTAGAG GCAGATCAGG TAAGCATACA TAGTAGAGGG AAAGGAGAGA
2301 ATGGAAATAA AACTGAATAT TATGCAGATT TATGCCTTAT TTTTTCAGAT
2351 TTTTAAAGGT TGGGTCTTTC AGGCTGGTTT TGGTTGTAT TAGATCTGTA
2401 TAGTTTAGTG ATTTAGTTT ATATTAAAGC TACGATTAAT ATTTTCTTCT
2451 TGGCGATATT TCTTTGCTTT TTTTAAAAA CAACCTTCCA TTTTTCAGATG
```



```

2501 TTTCGTTGAA TCTATTTAGA GCTTCACCAT GGCAATATGT ATTTCCCTTA
2551 AAACACTGCA AACAAATATA CTAGGAGTGT GCCCTTTTAA TCTTTACTAG
2601 TTATTGTGAG ACTGCTGTGT AAGCTAATAA ACACATTTGT AAAAACATTG
2651 TTTCGAGGAA GAAAACTTCG AGTTACAGGT CAGGAAAAGC CTGCTGAATT
2701 TATGTTGTAA ACGTTACTTA ACACAGTATA AAGATGAAAA GACAACAAAA
2751 GTATCTTCAT ACTTCCTCAT CCCCTCATTG CAACAAAACC TTAAACTGGG
2801 AGAACCTTAG TCCCCTCTCT TTCCTCTTCC TCCTCCACTT CCCACTTATT
2851 GCCACTTTGT AATATTCAGA GAGCACTTGG ATTATGGATC TGAATAGAGA
2901 AATGCTTACA GATAATCATT AGCCACATA CCAGTAACTT ATACTTAAAG
2951 ATGGGATGGA GTTATAAAGT GCTTTTATAA TCCAATATAA TTGCTAAAGG
3001 CAAGGGTTGA CTCITTGTTT TATTTTGACA TGGCATGTCC TGAATAAAT
3051 ATTGTTTAC TATGAAAAA AAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

98382468:

Rab proteins.

97203146:

GTP-bound forms of rab6 induce the redistribution of Golgi proteins into the endoplasmic reticulum.

Peptide information for frame 3

ORF from 456 bp to 1217 bp; peptide length: 254
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: BACTERIAL_OPSIN_RET (45-57)

```

1 MSAGGDFGNP LRFKLVFLG EQSVAKTSLI TRFRYDSFDN TYQAIIGIDE
51 LSKTMYLEDG TIGLRLWDTA QERLRSLIP RYIRDSAAAV VVYDITNVNS
101 FQQTWKIDD VTERGSDVI ITLVGNRTDL ADKQVSVEE GERKAKGLNV
151 TFIETRAKTG YNVKQLFRRV AAALPGMEST QDGSREDMSD IKLEKPQEQT
201 VSEGGCSCYS PMSSSTLPQK PPYSFIDCSV NIGLNLFP SL ITFCNSSLLP
251 VSWR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfk2_4k14, frame 3

PIR:G34323 GTP-binding protein Rab6 - human, N = 1, Score = 944, P = 6.5e-95

TREMBL:CET25G12_2 gene: "T25G12.4"; Caenorhabditis elegans cosmid T25G12., N = 1, Score = 756, P = 5.4e-75

TREMBL:NTNTRAF_1 gene: "Nt-rab6"; Nicotiana tabacum SR1 Nt-rab6 mRNA, complete cds., N = 1, Score = 698, P = 7.6e-69

TREMBL:D84314_1 product: "rab6"; Drosophila melanogaster mRNA for rab6, complete cds., N = 1, Score = 836, P = 1.9e-83

PIR:T01588 small GTP-binding protein Fl6B22.10 - Arabidopsis thaliana, N = 1, Score = 704, P = 1.8e-69

>PIR:G34323 GTP-binding protein Rab6 - human
 Length = 208

HSPs:

Score = 944 (141.6 bits), Expect = 6.5e-95, P = 6.5e-95
 Identities = 186/208 (89%), Positives = 190/208 (91%)

Query: 1 MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFYDSFDNTYQAIIGIDFLSKTMYLEDG 60
 MS GGDGFGNPLRKFKLVFLGEQSV KTSILITRF YDSFDNTYQA IGIDFLSKTMYLED
 Sbjct: 1 MSTGGDFGNPLRKFKLVFLGEQSVGKTSILITRFMYDSFDNTYQATIGIDFLSKTMYLED 60

Query: 61 TIGRLRLWDTAGQERLRSIPRYIRDSAAAVVYDITNVNSFQQTTKWIDDVTERGSDVI 120
 T+ L+LWDTAGQER RSLIP YIRDS AVVVYDITNVNSFQQTTKWIDDVTERGSDVI
 Sbjct: 61 TVRLQLWDTAGQERFRSLIPSYIRDSTVAVVVYDITNVNSFQQTTKWIDDVTERGSDVI 120

Query: 121 ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST 180
 I LVGN+TDLADKRQVS+EEGERKAK LNV FIET AK GYNVQLFRRVAAALPGMEST
 Sbjct: 121 IMLVGNKTDLADKRQVSIEEGERKAKELNVMFIETSAKAGYNVQLFRRVAAALPGMEST 180

Query: 181 QDGSREDMSDIKLEKPEQTVSEGGCSC 208
 QD SREDM DIKLEKPEQ VSEGGCSC
 Sbjct: 181 QDRSREDMDIKLEKPEQPVSEGGCSC 208

Pedant information for DKFZphfd2_4kl4, frame 3

Report for DKFZphfd2_4kl4.3

[LENGTH] 254
 [MW] 28385.29
 [pI] 7.58
 [HOMOL] PIR:G34323 GTP-binding protein Rab6 - human 1e-102
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YLR262c]
 7e-60
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YLR262c] 7e-60
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YOR089c] 2e-33
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YOR089c] 2e-33
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YOR089c] 2e-33
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c]
 2e-33
 [FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
 YGL210w] 3e-28
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 8e-27
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w]
 8e-27
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YOR101w]
 2e-21
 [FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 2e-21
 [FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YOR101w]
 2e-21
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 2e-21
 [FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S.
 cerevisiae, YOR101w] 2e-21
 [FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YOR101w] 2e-21
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 6e-19
 [FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 6e-19
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 6e-19
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 6e-16
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 6e-16
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 6e-16
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 4e-13
 [FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 4e-13
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 2e-09
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-08
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YLR229c] 8e-08
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 1e-05
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR094w] 5e-05
 [BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
 [SCOP] dlas3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domain 1e-32
 [SCOP] dimhl_ 3.29.1.4.2 Rac1 [Human (Homo sapiens)] 2e-51
 [SCOP] d5p21_ 3.29.1.4.1 cH-p21 Ras protein [human (Homo sapiens)] 7e-53
 [SCOP] dlhura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) [human (Homo sapiens)] 1e-46
 [SCOP] dia2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) [Do] 6e-60
 [PIRKW] nucleus 2e-14
 [PIRKW] cell cycle control 5e-15
 [PIRKW] membrane trafficking 3e-71
 [PIRKW] endoplasmic reticulum 1e-29
 [PIRKW] phosphoprotein 1e-29
 [PIRKW] prenylated cysteine 2e-36
 [PIRKW] signal transduction 5e-15
 [PIRKW] transforming protein 5e-30
 [PIRKW] purine nucleotide binding 1e-28
 [PIRKW] alternative splicing 1e-18
 [PIRKW] P-loop 3e-71

[PIRKW] lipoprotein 2e-36
 [PIRKW] proto-oncogene 1e-20
 [PIRKW] methylated carboxyl end 1e-20
 [PIRKW] membrane protein 1e-29
 [PIRKW] GTP binding 3e-71
 [PIRKW] thiolester bond 1e-29
 [PIRKW] Golgi apparatus 1e-29
 [SUPFAM] ras transforming protein 1e-76
 [PROSITE] BACTERIAL_OPSIN_RET 1
 [PFAM] Ras family (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D

SEQ MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDG
 1kao-CCEEEEEECTTTTCHHHHHHHHHHCCCCCCTTTTC-EEEEEEEEETTE

SEQ TIGLRLWDTAGQERLRLSLIPRYIRDSAAAVVYDITNVNSFQQTTKWIDDVTERGSDVI
 1kao- EEEEEEEECTTTTCHHHHHHHHHHCCCCCEEEETTTTHHHHHHHHHHHHHHHHHHTTCC

SEQ ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST
 1kao- EEEEEETTTTGGGCCCHHHHHHHHHHCCCCCEECTTTTHHHHHHHHHHH.....

SEQ QDGSREDMSDIKLEKPQEQTVEGGCSCYSPMSSSTLPQKPPYSFIDCSVNIGLNLFPSL
 1kao-

SEQ ITFCNSSLLPVSWR
 1kao-

Prosite for DKFZphfd2_4k14.3

PS00327 45->57 BACTERIAL_OPSIN_RET PDOC00291

Pfam for DKFZphfd2_4k14.3

HMM_NAME Ras family (contains ATP/GTP binding P-loop)

HMM *KLVIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK
 KLV++G+ +V K++L RF +++F++ Y + IG+DF++KT++++ TI 63

Query 15 KLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDGTIG

HMM LQIWDTAGQERYRSMRPMYYRGAMGFMVYDITNRqSFENIrNWweEIrR
 L +WDTAGQER RS+ P Y+R++ ++++VYDITN SF+ ++W++++R+ 113

Query 64 LRLWDTAGQERLRLSLIPRYIRDSAAAVVYDITNVNSFQQTTKWIDDVRT

HMM HCDrDENVPIMLVGNKCDLEDQQRQVstEEGQeFAREWGAIPFMETSAKTN
 + ++V+I LVGN +DL+D+RQVS EEG+ A+ ++ + F+ET AKT+ 160

Query 114 ERG--SDVIITLVGNRTDLADKRQVSVEEGERKAKGLN-VTFIETRAKTG

HMM iNVEEAFMEIvReIlqrMqe.q.NqteNinidQpsrnrk....rCCCIM*
 +NV++ F +++ +++ +++ + ++++++I+ +++++ + +C+ + 208

Query 161 YNVKQLFRRVAAALPGMESTQDGSREDMSDIKLEKPQEQTVEGGCS-C

DKFZphfkd2_4m11

group: transmembrane protein

DKFZphfbr2-4m11 encodes a novel 159 amino acid protein with weak similarity to the putative membrane protein YMR034c of *S. cerevisiae*.

The novel protein contains 4 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker of neuronal cells.

weak similarity to YMR034c

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1749 bp

Poly A stretch at pos. 1727, polyadenylation signal at pos. 1713

```

1 GGGGTCCTCA AAGCCGCCGG AGCAACCCCC AGGTCTTTAC TTTACAATCG
51 GCAATTGAC TTGCTCTGCT GCATGCTGCG AGGGACCAAG GAAAGTGTGG
101 AGACGCTCCA AGGATTAGGT GATCGGAGCT TGAAAAGAAA AAAAGCCAAA
151 CAAATAAACA AAACCCACCC ACCCTAACGA ATATGAGGCT GCTGGAGAGA
201 ATGAGGAAAG ACTGGTTCAT GGTGGAATA GTGCTGGCGA TCGCTGGAGC
251 TAAACTGGAG CCGTCCATAG GGTGAATGG GGGACCACTG AAGCCAGAAA
301 TAACTGTATC CTACATTGCT GTTGCAACAA TATTCTTTAA CAGTGGACTA
351 TCATTGAAAA CAGAGGAGCT GACCAAGTCT TTGGTGCATC TAAAACTGCA
401 TCTTTTATT CAGATCTTTA CTCTTGCAAT CTTCACAGCA ACAATATGGC
451 TTTTCTTCA GCTTTTATCA ATCACACCCA TCAACGAATG GCTTTTAAAA
501 GGTTCGAGA CAGTAGGTG CATGCCTCCG CCTGTGTCTT CTGCAGTGAT
551 TTTAACCAAG GCAGTTGGTG GAAATGAGGC AGCTGCAATA TTTAATTCAG
601 CCTTTGGAAG TTTTGGTA AGTAAACATA GTTTAACTTG TCTATTACAA
651 CTTTGTCTGT GATATTGTGT ATATGAAAGA TTTAGTGAAA GCTGGATTGT
701 TTTTACTCTT TGGTTAAGTA TAAAAATTGT TGAATCTTTT CATGTGCCAG
751 TATCCATACC CTGAAGAAA GTAGTTAATG AATAAGCAA ATGTTCTCTT
801 ACAATATATT TTGGAGGTTT GGATTTTAAA ATTCCATTTA ATGAATTCAA
851 GGAATCAATT AAAACACTAT GTGTCTCCTT ATAGAGGTTA TGTCAATATA
901 TTGATCATTT AATGAGGCTT TTAGATTAT TATTATTTTG TATCATGGGA
951 CTGAGGATTT TGAAAAGGAA ACATGACCCA GCTGGTCAGA AAGGGAATGC
1001 TAATTTACTT GTTGACATGC CATTATTTT GTACATTTC CTGTCAAAGA
1051 AGCTACTGGC TTGGATGCTT CTGAGAAATC TATGTGAGAA AAAATTTGAA
1101 AGGAAGATAT GACTAATGAG TAATTGCAA GTAAATGTTG TATCTATATA
1151 TATATATATA TAAAGATTCA AAAGTAGTTC AGCTTTCATA AGTAGAACCA
1201 ATATAAGGAC GTTGTTTTAG CATTTTTAA TATTATTTT AAATAAATGA
1251 TGTAACAGAG GCTTGATTG TGTTATGAAA GATTGAGAAA CTAAATTTTC
1301 TGTTGATTTA ATTTTGTGT GCCTTAAAC TTGTGTAAT TCCTGAAGTT
1351 AATTATCATA TTGTACTTTT TGGGGCATAA CTCATTAGCA GATATGTAGT
1401 GCAGTGATTT ACAATAATT GAGAGTAAAA TCAGTGATGT ATAACTAGT
1451 TCATGAGTCT AGGTAAAATA TCAATTACCT CTGTTTAAAA TGCTCTGTTA
1501 ATTATTATTG TATGATTTA AATGTAGTTA AAGCTTTTAA ACATGTTGTT
1551 ACATAGTGTT AATTCTACAC AGTGCTACAC AGCTTTTAGT GTCACATAGC
1601 CTTACAGAGT TTATAATGAT GTAGCATCTG CAAAATATAT GCATAGCTTA
1651 TATCCTATTT TTATAGAGCC AGTAATGGTT TTTGTGATGC TGTATTACTT
1701 CTGGGTTTAA GACAATAAAG TCTGTTTAA CAAAAAATAA AAAAAAATAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3


```

1  MRLLEMRKD  WFMVGIVLAI  AGAKLEPSIG  VNGGPLKPEI  TVSYIAVATI
51 FFNGLSLKT  EELTSALVHL  KHLHFIQIFT  LAFFPATIWL  FLQLLSITPI
101 NEWLLKGLQT  VGCMPPPVSS  AVILTAVGG  NEAAAFINSA  FGSFLVSKHS
151 LTCLQLLLL

```

BLASTP hits

Alert BLASTP hits for DKFZphfkd2_4m11, frame 3

PIR:A65015 yfeH protein - Escherichia coli (strain K-12), N = 1, Score = 131, P = 4.2e-08

HSPs:

Query: 5 ERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPKLPEITVSYIAVATIFFNSGLSLKTEELT 64
E ++ WF + + + I A+ P+ +GG +K + ++ Y VA IF SGL +K+ L
Sbjct: 18 EFLKSQWFFICLAILIVIRFAPNFARDDGGLIKGQYSIGYGCVAWIFLQSLGLMKRSRLM 77

Query: 65 SALVHLKHLFLFIQFTLAFFPATIWLF---LQLLSITPINEWLLKGLQTVGCMPPPVSSA 121
+ + + + H I + + + + F + + + I + W L G L P V + S
Sbjct: 78 ANMLNWRHAHATILVLSFLITSSIVYGCCAVKAANDPKIDDWVLIGLILTATCPTTVASN 137

```
Query: 122 VILTKAVGGNEAAAFNSAFGSFL 145
      VI+T  GGN      +      G+ L
Sbjct: 138 VIMTTNAGGNSLLCVCEVFIGNLL 161
```

Pedant information for DKF2phfkd2_4m11, frame 3

Report for DKFZphfkd2 4m11.3

```
[LENGTH]      159
[MW]           17282.92
[pI]           9.06
[HOMOL]        PIR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces cerevisiae)
5e-12
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YMR034c] 2e-13
[PROSITE]      MYRISTYL 2
[PROSITE]      PKC_PHOSPHO_SITE 1
[PROSITE]      TRANSMEMBRANE 4
[KW]           4
```

```

SEQ      MRLLEMRKRDWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSSGLSLKT
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccchhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM..

```

```
SEQ      EELTSALVHLKHLFLFIQIFTLAFFPATIWLFLQLLSITPINELLKGLQTVGCMPPPVSS
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhcccccchhhhhhhheeeccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....
```

```

SEQ      AVILTKAVGGNEAAAI FNSAFGSLVSKHSLTCLLQLLL
PRD      ceeeeccccchhhhhhccccceeeceeeeeeeccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMM

```

Prosites for DKFZphfkd2 4m11.3

PS00005	57->60	PKC_PHOSPHO_SITE	PDOC00005
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	129->135	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfk2_4m11.3)

PAGE INTENTIONALLY LEFT BLANK

DKFZphutel_17k7

group: uterus derived

DKFZphutel_17k7 encodes a novel 520 amino acid protein with weak similarity to *S. Cerevisiae* Fip1.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to *S.cerevisiae* Fip1

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1914 bp

Poly A stretch at pos. 1897, polyadenylation signal at pos. 1867

```
1 CGGACGCGTG GCGGACGCG TGGGGCCTTC CTGGGATTGG AGTCTCGAGC
51 TTTCTTCGTT CGTTCGCCGG CGGGTTTCGG CCCTTCTCGC GCCTCGGGGC
101 TCGGAGGCTG GGAAGGGGT TGGAGGGGGC TGTGATCGC CGCGTTTAAG
151 TTGCGCTCGG GCGGCCCATG TCGGCCGGCG AGGTCGAGCG CCTAGTGTCTG
201 GAGCTGAGCG GCGGGACCGG AGGGGATGAG GAGGAAGAGT GGCTCTATGG
251 CGATGAAAAT GAAGTTGAAA GGCAGAGAAG AGAAAATGCC AGTGCTAATC
301 CTCCATCTGG AATTGAAGAT GAACTGCTG AAAATGGTGT ACCAAAACCG
351 AAAGTGACTG AGACCGAAGA TGATAGTGAT AGTGACAGCG ATGATGATGA
401 AGATGATGTT CATGTCACCTA TAGGAGACAT TAAACCGGGA GCACCACAGT
451 ATGGGAGTTA TGGTACAGCA CCTGTAAATC TTAACATCAA GACAGGGGGA
501 AGAGTTTATG GAACTACAGG GACAAAAGTC AAAGGAGTAG ACCTTGATGC
551 ACCTGGAAGC ATTAATGGAG TTCCACTCTT AGAGGTAGAT TTGGATTCTT
601 TTGAAGATAA ACCATGGCGT AACCTGGTGT CTGATCTTTC TGATTATTTT
651 AATTATGGGT TTAATGAAGA TACCTGGAAG GCTTACTGTG AAAAACAAAA
701 GAGGATACGA ATGGGACTTG AAGTTATACC AGTAACCTCT ACTACAAATA
751 AAATTACGGT ACAGCAGGGA AGAAGTGGAA ACTCAGAGAA AGAACTGCC
801 CTCCATCTTA CAAAGCTGA GTTACTTCTT CCTCCTTCTT TGTTCAAGAC
851 TGGGCTTCCA CCGAGCAGGA GATTACCTGG GCAATTGAT GTTATCGGTC
901 AGACTATAAC TATCAGCCGA GTAGAAGGCA GGCAGCGGCG AAATGAGAAC
951 AGCAACATCA AGGTCCTTTC TGAAAGATCT GCTACTGAAG TAGACAACAA
1001 TTTTAGCAAA CCACCTCCGT TTTTCCCTCC AGGAGCTCCT CCCACTCACC
1051 TTCCACCTCC TCCATTTCTT CCACCTCCTC CGACTGTCTG CACTGCTCCA
1101 CCTCTGATTC CACCACCGGG TTTTCCCTCT CCACCAGGCG CTCCACCTCC
1151 ATCTCTTATA CCAACAATAG AAAGTGGACA TTCCTCTGGT TATGATAGTC
1201 GTTCTGCACG TGCATTTCCA TATGGCAATG TTGCCTTTCC CCATCTTCCT
1251 GGTCTTGCTC CTTCGTGGCC TAGTCTTGTG GACACCAGCA AGCAGTGGGA
1301 CTATTATGCC AGAAGAGAGA AAGACCGAGA TAGAGAGAGA GACAGAGACA
1351 GAGAGCGAGA CCGTGATCGG GACAGAGAAA GAGAACGCAC CAGAGAGAGA
1401 GAGAGGGAGC GTGATCACAG TCCTACACCA AGTGTTTCA ACAGCGATGA
1451 AGAACGATAC AGATACAGGG AATATGCAGA AAGAGGTAT GAGCGTCACA
1501 GAGCAAGTCG AGAAAAAGAA GAACGACATA GAGAAAGACG ACACAGGGAG
1551 AAAGAGGAAA CCAGACATAA GTCCTCTCGA AGTAATAGTA GACGTCGCCA
1601 TGAAAGTGAA GAAGGAGATA GTCACAGGAG ACACAACAC AAAAAATCTA
1651 AAAGAAGCAA AGAAGGAAAA GAAGCGGGCA GTGAGCCTGC CCCTGAACAG
1701 GAGAGCACCG AAGCTACACC TGCAGAATAG GCATGGTTT GGCCTTTTGT
1751 GTATATTAGT ACCAGAAGTA GATACTATAA ATCTTGTTAT TTTTCTGGAT
1801 AATGTTTAAG AAATTTACCT TAAATCTTGT TCTGTTTGT AGTATGAAAA
1851 GTTAACTTTT TTTCCAAAAT AAAAGAGTGA ATTTTTCATG TTAAGTTAAA
1901 AAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 168 bp to 1727 bp; peptide length: 520
 Category: similarity to known protein

```

1 MSAGEVERLV SELSGGTGGD EEEEWLYGDE NEVERPEEN ASANPPSGIE
51 DETAENGVPK PKVTETEDDS DSDDDDDD VHTIGDIKT GAPQYGSYGT
101 APVNLNIKTG GRVYGTGK VKGVLDAPG SINGVPLEV DLDSFEDKEW
151 RKPGADLSY FNYGFNEDTW KAYCEKQRI RMGLEVIPVT STTNKITVQQ
201 GRTGNSEKET ALPSTKAEFT SPPSLFKTGL PPSRLPGAI DVIQTITIS
251 RVEGRRRANE NSNIQVLSER SATEVDNNS KPPFFPPGA PPTHLPPIPF
301 LPPPTVSTA PPLIPPPGFP PPGAPPPSL IPTIESGHSS GYDSRSARAF
351 PYGNVAFPHL PGSAPSWPSL VDTSKQWYY ARREKDRDRE RDRDRDRDR
401 RDRERERTRE RERERDHSPT PSVNSDEER YRYREYAERG YERHRASREK
451 EERHRERRHR EKEETRHKSS RSNSRRRHES EGDShRRHK HKKSKRSKEG
501 KEAGSEPAPE QESTEATPAE

```

BLASTP hits

Entry AF016427_4 from database TREMBL:

gene: "F32D1.9"; Caenorhabditis elegans cosmid F32D1.

Score = 392, P = 1.8e-36, identities = 156/519, positives = 212/519

Entry S62454 from database PIR:

hypothetical protein SPAC22G7.10 - fission yeast (Schizosaccharomyces pombe)

Score = 246, P = 2.0e-22, identities = 62/163, positives = 91/163

Entry A56545 from database PIR:

FIP1 protein - yeast (Saccharomyces cerevisiae)

Score = 186, P = 2.9e-16, identities = 56/206, positives = 92/206

Alert BLASTP hits for DKFZphut1_17k7, frame 3

TREMBLNEW:AF109907_1 product: "S164"; Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds., N = 2, Score = 236, P = 1.5e-16

>TREMBLNEW:AF109907_1 product: "S164"; Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds.

Length = 735

HSPs:

Score = 236 (35.4 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
 Identities = 51/120 (42%), Positives = 76/120 (63%)

```

Query:   383 REKDRDRERDRDRDRDRDRERERTREERERERDHSPTPSVNSDEERYRYREYA---ER 439
          REK+++RER+R+RDRDR +ER+R R+RER+RD S + +++R R RE + ER
Sbjct:   227 REKEKERERERERDRDRDRTKERDRDRDRDRDRDRERS--DRNKDRSRSREKSRDRER 285

```

```

Query:   440 GYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHSEEGDShRRHKHKKSKRSK 498
          ER R + ER RER R RE+E R + + +R E +E D++ R K ++ R K
Sbjct:   286 EREREREREREREREREREREREREREREREKDKKRDREDEEDAYERRKLEKLEK 345

```

Query: 499 E 499

Sbjct: 346 E 346

Score = 214 (32.1 bits), Expect = 4.4e-14, Sum P(2) = 4.4e-14
 Identities = 50/133 (37%), Positives = 75/133 (56%)

```

Query:   383 REKDRDR-ERDRDRERDRDRERERTREERERERDHSPTPSVNS-DEERYRYREYAERG 440
          RE++R+R ER+R+RER+R+R++E+ER RERER+RD T D ER R R+ ER
Sbjct:   208 REREREREREREREREREREKEKERERERERDRDRDRTKERDRDRDRDRDRD-REERS 266

```

```

Query:   441 YERHRASREKEERHRERHRREKEETRHKSSRSNSRRRHSEEGDShRRHKHKKSKRSKEG 500
          +R++ E+ R+R RE+E R + R R R E + R + ++ K K
Sbjct:   267 SDRNKDRSRSREKSRDRE-RERERERE-REREREREREREREREREREKDKKRD 324

```

Query: 501 KEAGSEPAPEQESTE 515

+E E A E+ E

Sbjct: 325 REEDEDAYERKLE 339

Pedant information for DKFZphutet1_17k7, frame 3

Report for DKFZphute1_17k7.3

```

SEQ      MSAGEVERLVSELSGGTGGDEEEEWLYGDENEVERPEEENASANPPSGIEDETAENGVPK
SEG
PRD      cccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PKVTETEDSDSDSDDDDDDVHVTIGDIKTGAPQYGSYGTPAVNLNIKTGGRVYGTGTK
SEG      .. xxxxxxxxxxxxxxxxxxxxx .....
PRD      cceeeccccccccccccccccceeeccccccccccccccccccccceeecccccecccccc

SEQ      VKGVLDLAPGSINGVPLEVDLDSFEDKPWRKPGADLSDYFNYGFNEDTWKAYCEKQKRI
SEG
PRD      ceeeccccccccccccceeeccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ      RMGLEVIPVTSTTNKITVQQGRGTNSEKETALPSTKAFTSPPSLFKTGLPSPRRLPGA
SEG

```



```

PRD      hhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      DVIGQTITISRVEGRRRANENSNIQVLSERSATEVDNNSKPPFFPPGAPTHLPPPPF
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      LPPPTVSTAPLIPPGFPPPPGAPPPSLIPTIESGHSSGYDSRSARAFYPGNVAFPHL
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PGSAPSWPSLVDTSKQWDYYARREKDRDRERDRDRDRERERERERERERERERDHSPT
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcccccc

SEQ      PSVFNSDEERYRYREYAERGYERHRASREKEERHRERHREKEETRHKSSRSNSRRRHES
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccchhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc

SEQ      EEGDSHRRHKKHKKSKRSKEGKEAGSEPAPEQESTEATPAE
SEG      xx..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphut1_17k7.3

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	278->282	ASN_GLYCOSYLATION	PDOC00001
PS00005	169->172	PKC_PHOSPHO_SITE	PDOC00005
PS00005	193->196	PKC_PHOSPHO_SITE	PDOC00005
PS00005	206->209	PKC_PHOSPHO_SITE	PDOC00005
PS00005	214->217	PKC_PHOSPHO_SITE	PDOC00005
PS00005	233->236	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	469->472	PKC_PHOSPHO_SITE	PDOC00005
PS00005	474->477	PKC_PHOSPHO_SITE	PDOC00005
PS00005	485->488	PKC_PHOSPHO_SITE	PDOC00005
PS00005	494->497	PKC_PHOSPHO_SITE	PDOC00005
PS00006	2->6	CK2_PHOSPHO_SITE	PDOC00006
PS00006	17->21	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	426->430	CK2_PHOSPHO_SITE	PDOC00006
PS00007	434->442	TYR_PHOSPHO_SITE	PDOC00007
PS00007	152->161	TYR_PHOSPHO_SITE	PDOC00007
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	154->160	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS00008	244->250	MYRISTYL	PDOC00008
PS00008	289->295	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00009	253->257	AMIDATION	PDOC00009

(No Pfam data available for DKFZphut1_17k7.3)

DKFZphut1_18c12

group: uterus derived

DKFZphut1_18c12 encodes a novel 378 amino acid protein nearly identical to human WUGSC:H_DJ0872F07.1 protein.

The novel protein has an additional N-terminal domain, which is not present in WUGSC:H_DJ0872F07.1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

nearly identical to human WUGSC:H_DJ0872F07.1 protein

on genomic level encoded by AC004537, 10 exons the predicted protein sequence AC004537_1 is only partially o.k. first exon wasn't predicted there are additional exons predicted (BLASTX/EST-BLAST shows that the cDNA is only partly spliced) intron ~1216-3540/-3577-5059

Sequenced by AGOWA

Locus: map="7q31"

Insert length: 6005 bp

Poly A stretch at pos. 5980, polyadenylation signal at pos. 5968

```
1  ACGGGGTGCT  GCTAGCGGAG  GCGCCATATT  GGAGGGGACA  AAACCTCCGGC
51  GACAGCGAGT  GACACAAATA  AACCCCTGGA  CCCCTTGTT  CCCTCAGCTC
101 TAAGGGCCGC  GATGTTGTAC  CTAGAAGACT  ATCTGGAAAT  GATTGAGCAG
151 CTTCTATGG  ATCTGCGGGA  CCGCTTCACG  GAAATGCGCG  AGATGGACCT
201 GCAGGTGCAG  AATGCAATGG  ATCAACTAGA  ACAAAGAGTC  AGTGAATTCT
251 TTATGAATGC  AAGAAAAAAT  AAACCTGAGT  GGAGGGAAGA  GCAAATGGCA
301 TCCATCAAAA  AAGACTACTA  TAAAGCTTTG  GAAGATGCAG  ATGAGAAGGT
351 TCAGTTGGCA  AACCAGATAT  ATGACTTGGT  AGATCGACAC  TTGAGAAAGC
401 TGGATCAGGA  ACTGGCTAAG  TTTAAATGG  AGCTGGAAGC  TGATAATGCT
451 GGAATTACAG  AAATATTAGA  GAGGCGATCT  TTGGAATTAG  ACACTCCTTC
501 ACAGCCAGTG  AACAATCACC  ATGCTCATTC  ACATACTCCA  GTGGAAAAAA
551 GGAATATATA  TCCAACCTCT  CACCATACGA  CAACAGATCA  TATTCCTGAA
601 AAGAAATTTA  AATCTGAAGC  TCTTCTATCC  ACCCTTACGT  CAGATGCCTC
651 TAAGGAAAAA  AACTAGGTT  GTCGAAATAA  TAATTCACAC  GCCTCTCTA
701 ACAATGCCTA  CAATGTGAAT  TCCTCCCAAC  CTCTGGGATC  CTATAACATT
751 GGCTCGTTAT  CTTCAGGAAC  TGGTGCAGGG  GCAATTACCA  TGGCAGCTGC
801 TCAAGCAGTT  CAGGCTACAG  CTCAGATGAA  GGAGGGACGA  AGAACATCAA
851 GTTTAAAAGC  CAGTTATGAA  GCATTTAAGA  ATAATGACTT  TCAGTTGGGA
901 AAGAATTTT  CAATGGCCAG  GGAACAGTT  GGCTATTTCAT  CATCTTCGGC
951 ACTTATGACA  ACATTAACAC  AGAATGCCAG  TTCATCAGCA  GCCGACTCAC
1001 GGAGTGGTCG  AAAGAGCAAA  AACAACAACA  AGTCTTCAAG  CCAGCAGTCA
1051 TCATCTTCCT  CCTCCTCTTC  TTCCTTATCA  TCGTGTTCCT  CATCATCAAC
1101 TGTGTACAA  GAAATCTCTC  AACAAACAAC  TGTAGTGCCA  GAATCTGATT
1151 CAAATAGTCA  GGTGATTGG  ACTTACGACC  CAAATGAACC  TCGATACTGC
1201 ATTTGTAATC  AGGTAAAAGT  CTGTTATATC  TATAAAAGTA  TAATCTGAAT
1251 AAACTAGAAG  GAAGAGAACT  ATTTCAATTT  TAAGCACTTT  TTTAAACTCA
1301 CTTAAAATAC  CTTTGCTTTA  TTTGTATACT  TTTCTCCCC  TTCTTACAAA
1351 AGTGACATTT  GCTGTAAATA  CTGAGTATAA  AGAAAAATGT  TACCCATAAT
1401 CCTAGCCCTC  AGATACAACC  TGTAACATAA  CATTTTGGT  ATACCACTAC
1451 CATATACCTC  ATGTGCACAT  TGGCTGCCTT  AATAAAATAC  AACAGACTGG
1501 GTAGCTTAAA  CAACAGAAAA  TAATTTTCTC  ACAGGTATGA  AGGCTGGGAA
1551 GTCCAAGATC  AAGGTGTCCA  CTGACTCAGT  TCTGGAGGAG  GGCTCCCTTC
1601 CTAGATGGAG  ACTGCTGCCT  TCTCACCAGG  TCCTCACATG  ATAGAGGGAG
1651 AAAGAGTGTG  CTCTGGTGTC  TTTTCTTATA  AGGGCACCAG  CCTTGTGAGA
1701 GTAGGACCCC  ACTCTATGAC  CTCATTTAAC  CTTTACCACC  TCCTCACAGG
1751 CCCTGTTTCC  AATTATAGTC  ACGTTGGGGG  TTAGGGCTTC  AACATATGAT
1801 TTTGAGACAT  AAGCTTGCAT  TTCATAACAC  GTGTCTATGC  AGATTGTCAC
1851 ATGCATGTGT  GTATAAGTTT  GTCAGTAGGA  ACCACAGTGT  ATACTTTCTT
1901 GTTACTGGCT  TTTTCTCTA  AATCAGGTAT  ACCGAACATG  ATTTTCTTTT
1951 AAGATCATAT  TTTTAATTTT  CACATAGTTA  TCTCTTATGC  CATCCAGTGT
2001 AGTTTCTTTA  ACCAATACCT  AGCTATAGAT  TATATTAGTG  GTTTTAATTT
2051 GTTTGAAATT  AGGGATAATA  TTACGATAGG  CATTTTTTAA  ATGTAATCCA
2101 TTTTATACAT  CTAATTTCTT  GGATAATCTT  TTAGAAATAA  AATTAGGCTG
2151 TAAATATTTG  ACAGACACCA  AAATATATTT  TCTAGAAATT  TATTACCAAA
2201 AATTAAATAA  CATACCGGTT  TACTAAACCC  TGTCCAACAC  TGGATATTAT
2251 TTTCTTTTAA  AAACCTAAGT  CCAATTTGGT  AGTTTATAT  TATGATTGTT
2301 TTAATACAC  TAGTATTATT  GAAGTTGGAC  ATTTTGTGAC  CATTTTGTGT
2351 TTTACATTA  TGAATCGACT  CCTAATGGTG  TCGGCTGATT  TTTCTATTGT
```



```

2401 TTTTGTATG TACTCTAAAT ATTTGCTTGA TTTAGTTTTT TAAAAATAAT
2451 TCTAAAAATT TAATTTTATG TAGTTATGAC TGTTAATTTT TTTTATGAA
2501 GCAAGCCATG GATTATATAC TTAGAAGGGC TTCTCTTTTG GCTCTTCTTT
2551 CTACAAAAAA TTGTCTTGTA TAATATTTTC TCCTAGTTTT TATATGGTTT
2601 TGTCTAGTTC TTTGCATGCT TCAGTTTCTT CACATTTAAG ACTTAGTCTA
2651 TCAGCAGATT ATTGTGTCTA ACAGTATGAG TTGCCAGTCT GATTTTAA
2701 AATTTTAACA ATTTGTTAGC TGTTCCTACTA TCACCCGATA AACATTTTTC
2751 AGTACAAATG ATAGAAAAGC ATATCCTGTA TCCTGACAAC AAAAGTAGAT
2801 TACTTGCAAA AGAACAAAAT CAGACTGAAC CTAGAGTTTT CCTCTGTAAC
2851 ACTAAAAAAC TAGAAGGTGA TGAATATGT CTGTAGAGCT TTCAGGGA
2901 AATTAAGAGC CCCCCAAAAC TTGATATTCA GAGAAGTTAT TTCTCTGCAT
2951 AGGACCATGT AAATATATTT TCACTCATGC AGAGAATCAG AAGATATTGCC
3001 ATCTAGTTAA TCCTGTCTGA AAAATTATTC AATCCACTGA GAACCTCAGT
3051 GAATCAAGA ATTAGCAAGT TATGCCCTAA AGTGCTGGTG ATGAAGAGCA
3101 AAAGAAAAAT GAGAAAGGAC ATAAAAAGA TAAGTTTAGA AGTTTCAAGG
3151 AAGGAGACTA TTAATTGCAA AAATATATAT GACCTAATGT GACCCAAGAA
3201 GTAAAACTT TCAGTAAGTA AATAATCAAG AAAGGAACCT AAAATTTTAA
3251 CAATAAGAAC TACCCAGAAA GATGACTCCT TCATCCGGGT GATTTATATG
3301 TCAAGTTCTT CCAGACTTCT GAAGGGCAGA TAATTCCTGT GCATTTCTTC
3351 CCACCCTTGC CCCACCCTGC CCAAAAGAGT ATTTTCAGAA AAAATTATTA
3401 TACCTTGATT CTCAATGTAA TTGTATATTC AGTGTATTTT CCTTTATTTT
3451 CCAGCAGTAT CATACATAAA CAGTTAATTG GTATCTAGGT GTTTGTACA
3501 TAGTCATAAT AAAGACATTT AATTTTTTTT AACTAGGTAT CTTATGGTGA
3551 GATGGTGGGA TGTGATAACC AAGATGTAAG TATTACATTT TTCTATTAG
3601 GAATGAAAAA AATCACAGGT TGTATTACT TGAATATTG TCTTATTGTC
3651 TGTATGGTTT GGTCTAAGAA AACAGGTTTG CAGGTATATT AGTTATGTTA
3701 TGCTAATGCT AGAATATTCC TCTTCAAAAT AGGGTAGTGT CCCTTAATGT
3751 GTTCCCTATT TTAATTTTAA AAGCTAATTT TATGGTTTTA TGTGCAGATT
3801 GTCTCAGAAG TGTATGTTG TATGAAAATT ATAAATACCC TCCTTTCCCT
3851 TTAATAAAAA ATACTGTGTT TACTAGAATC CAGTTCATTT ATCACATTGA
3901 AGAAATGGAA TTTTAAACA ATTCATTCTT TCAGGCTGCA CCGTGCTAAA
3951 GTGAAGGGTG GGATAATTGA GGATCTAATG TGAGATTATC TTCTCTCAT
4001 GAGTATAATA TTTTTCCTG TACTCTGCAG GTGTCAGCTG ATAAGAGCCA
4051 CCCCTGATCT AAAAAGTAAA GGAATTTGA AAGGAAGGAA TTCTTGGTTT
4101 TTAGGAGACT TAATTTTAGT TAGAGATACG TTTTATTATC AATACTGAGA
4151 ATATTGTTGT CTAGTAATTT TGACTCCCTC CTTATTAGT AGTGACAGGA
4201 TCCTAAGATT AACAGAGTT TTAATTTGT AAAACAATCT GAAGATTGAG
4251 GGAGCTGGCT AGGTGCATTA AAATGTGTAC TTTTCTAGA CCTGATAGGG
4301 TTACAGCAAC ATGCTCACGT AGATTGGGAC AGAGCCTCCT TCTGTTTCCC
4351 TGTCTAGATT CCCTGTAGG CTGTTTGTGG TTGTTGCAAA AACAATATTG
4401 CCAACCAATT TCAAGACAT CACTGTAAAC TCTTCTGGGG CAGTTAGTGA
4451 AAATGATGAA TGAGATTCTT ATGAGTACCA GCATCATGCT TCTCTGATTC
4501 TTCTTATTCC CAGTTGTGCT CTTCTGAGTG CTAAGACTTT CATGAAAGAG
4551 TTTTCTGCTT AATATGTTT AAAGAGGAAT AATTTTCTC TACATTTCAA
4601 GGAATAGAAA CACCCACGTA GGAATGCAG GGCATAAGAC ATAAATTAAT
4651 GTCTTTAATT ACAATCAGCT TATTCTACTT TATGAGACAG CAAATAAGGC
4701 TGACTATTAA ATAAAACTT AAGTTATATT TACCTTCTAC ATAGAAGATT
4751 CATCCCACTT CTTTTTGCCC TTGAAAGCTG AAAACTAGTG AATTTTCATT
4801 CATTAGGATG AGGGGACTAG ATTACATGGA CCTCAGGATT CTGGAAGATG
4851 CATAATTTTT CTGTGCCTTC ATTTCCCTCAT TCCTGAAGCT TATCATTTAG
4901 TCTAAATGAT GTCTAAATAA TCTAGATCTA AAAATTCTGA TGTACACAT
4951 CTAATTATTG TTAATTTAAA TGGATTATTC AGTCTCCTGA GCATATTTTA
5001 ATATACTCTC TTGTCTTCAG AAGTACTGAA AACTTGTTTT TTGCAATTTT
5051 GCTTTCTAGT GCCCTATAGA ATGGTTCCAT TATGGCTGCG TTGGATTGAC
5101 AGAGGCACCA AAGGCAAAAT GGTACTGTCC ACAGTGCAC TGTGCAATGA
5151 AGAGAAGAGG CAGCAGACAC AAATAAAGGT GGTCTTTTG TTTGATGAAG
5201 AAATAAACTT CAGCTGAAGA TTTTATATAG GACTTTAAAA AGAAGAGAAG
5251 AGAAGAAGA AACAAATGCAT TTCCAGGCAA CCACCTAAAG GATTACATA
5301 GACAATCCTA TAAGATCTTG AACTTGAATT TTATGGGTTG TATTTAATA
5351 ATGTAAGTAA ATTATTTATG CACTCCTGGT GTGCTATGAA TATTATTCCA
5401 GTTAGCCTTG GATTATTTCA GTGGCCAACA TATGCAGACA TTTGTAATCC
5451 TCAACCAATT TCTCAAAGTA ATGGGCATTG TATGATTTAG ACTTCAAGGA
5501 ATTCCAATGA TGAAGATTTT AAGGAAAGTA TTTTATATTC AACAGGTATA
5551 TTCTGCTGCA TGTACTGTAC TCCAGAGCTG TTATGTAAAC CTGTATATAA
5601 ATGGTTGCAA AAAAAAAAAA AAGTCAGTGC TTCTAAAAAG AATTTAAGAT
5651 AATGGTTTTT AAAATGCCCT TATAATAAGC TTTGTTTCTT TGTGAACTA
5701 AATCAGCAGG CTGAAGGAAA TGGTTTCATG GATAATGTGG GCTGGTATCC
5751 TCTAGAGTAC CTGGGTACAT AAACAGAAAC TCCTGTAGGT AAAAAGTAAT
5801 TTGTGCCATT AGTCTTCTA TGTCTTCTCA TCCAGATAGA GTGCAGTTCA
5851 TGAGGGAGGG GCGGGGGGAC TGAAGGGGAA AGGGCGTTAA AGTGATACAT
5901 TTTTATACCA AATGTGTTTA TTTTGTGTG CAAGTAATCC TTAATTTGTC
5951 AATTGTATTA GGTGTAAAA TAAAGTTTTT AAAAAATTA AAAAAAATAA
6001 AAAAA

```

BLAST Results

Entry HSG20547 from database EMBL:
HSG20547| human STS A005W09.
Length = 154

Minus Strand HSPs:
 Score = 770 (115.5 bits), Expect = 2.9e-26, P = 2.9e-26
 Identities = 154/154 (100%)

Medline entries

98101645:
 The candidate tumour suppressor p33ING1 cooperates with p53 in cell growth control.

Peptide information for frame 1

ORF from 112 bp to 1245 bp; peptide length: 378
 Category: similarity to known protein

```

1 MLYLEDYLEM IEQLPMDLRD RFTEMREMDL QVQNAMDQLE QRVSEFFMNA
51 KKNKPEWREE QMASIKKDY KALEDADEKV QLANQIYDLV DRHLRKLDQE
101 LAKFKMELEA DNAGITEILE RRSLELDTPS QPVNNHHAHS HTPVEKRKYN
151 PSHHTTTDH IPEKKFKSEA LLSTLTSDAS KENTLGCRNN NSTASSNNAY
201 NVNSSQPLGS YNIGSLSSGT GAGAITMAAA QAVQATAQMK EGRTSSSLKA
251 SYEAFKNNDF QLGKEFSMAR ETVGYSSSSA LMTTLTONAS SSAADSRSGR
301 KSKNNNKSSS QSSSSSSSSS SLSSCSSSST VVQEISQQT VVPESDSNSQ
351 VDWTYDPNEP RYCICNQVKV CYIYKSII

```

BLASTP hits

Entry AF044076_1 from database TREMBL:
 "ING1"; product: "candidate tumor suppressor p33ING1"; Homo sapiens candidate tumor suppressor p33ING1 (ING1) mRNA, complete cds. Homo sapiens (human)
 Length = 279
 Score = 162 (57.0 bits), Expect = 1.1e-09, P = 1.1e-09
 Identities = 48/183 (26%), Positives = 92/183 (50%)

Entry AC004537_1 from database TREMBL:
 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07 from 7q31, complete sequence.
 Score = 1814, P = 3.7e-187, identities = 358/358, positives = 358/358

Entry CEY51H1A_1 from database TREMBL:
 gene: "Y51H1A.4"; Caenorhabditis elegans cosmid Y51H1A
 Score = 213, P = 3.7e-15, identities = 37/123, positives = 82/123

Alert BLASTP hits for DKFZphut1_18c12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18c12, frame 1

Report for DKFZphut1_18c12.1

```

[LENGTH]      378
[MW]           42275.72
[pI]           5.72
[HOMOL]        TREMBL:AC004537_1 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07
from 7q31, complete sequence. 1e-157
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YHR090c] 8e-05
[UNCAT]        04.05.01.04 transcriptional control [S. cerevisiae, YNL097c] 2e-04
[PROSITE]      MYRISTYL 3
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 4
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 3
[PROSITE]      ASN_GLYCOSYLATION 5
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 20.63 %

```


Prosites for DKFZphutel1_18c12.1

(No Pfam data available for DKFZphute1_18c12.1)

DKFZphutel_18i19

group: transcription factors

DKFZphutel_18i19 encodes a novel 759 amino acid protein with similarity to the SREBP-2 mutant sterol regulatory element binding protein-2 of *Cricetulus griseus*.

The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH₂-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

The new protein can find application in modulating/blocking the expression of genes involved in lipid metabolism.

similarity to transcription factor SF3

complete cDNA, complete cds, EST hits
strong similarity to mutated SREBP-2 of hamster,
similarity is not to SREP-2 part of protein but to the unknown part of
the fusion protein

Sequenced by AGOWA

Locus: /map=12

Insert length: 3664 bp

Poly A stretch at pos. 3647, polyadenylation signal at pos. 3636

```
1 GCGCTAGGTA GAGCGCCGGG ACCTGTGACA GGGCTGGTAG CAGCGCAGAG
51 GAAAGGCGGC TTTAGCCAG GTATTTCAGT GTCTGTAGAC AAGATGGAAT
101 CATCTCCATT TAATAGACGG CAATGGACCT CACTATCATT GAGGGTAACA
151 GCCAAGAAGC TTTCTCTTGT CAACAAGAAC AAGTCATCGG CTATTGTGGA
201 AATATTCTCC AAGTACCAGA AAGCAGCTGA AGAAACAAAC ATGGAGAAGA
251 AGAGAAGTAA CACCGAAAT CTCTCCAGC ACTTAGAAA GGGGACCTTG
301 ACTGTGTTAA AGAAGAAGTG GGAGAACCCA GGGCTGGGAG CAGAGTCTCA
351 CACAGACTCT CTACGGAACA GCAGCACTGA GATTAGGCAC AGAGCAGACC
401 ATCCTCCTGC TGAAGTGACA AGCCACGCTG CTTCTGGAGC CAAAGCTGAC
451 CAAGAAGAAC AAATCCACCC CAGATCTAGA CTCAGGTCAC CTCCTGAAGC
501 CCTCGTTCAG GGTCGATATC CCCACATCAA GGACGGTGAG GATCTTAAAG
551 ACCACTCAAC AGAAAGTAAA AAAATGGAAA ATTGTCTAGG AGAATCCAGG
601 CATGAAGTAG AAAAATCAGA AATCAGTGAA AACACAGATG CTTGCGGCAA
651 AATAGAGAAA TATAATGTTT CGCTGAACAG GCTTAAGATG ATGTTTGAGA
701 AAGGTGAACC AACTCAAACCT AAGATTCTCC GGGCCCAAAG CCGAAGTGCA
751 AGTGGGAAGGA AGATCTCTGA AAACAGCTAT TCTCTAGATG ACCTGGAAAT
801 AGGCCCAGGT CAGTTGTCAT CTTCTACATT TGACTCGGAG AAAAATGAGA
851 GTAGACGAAA TCTGGAACCT CCACGCCCTCT CAGAAACCTC TATAAAGGAT
901 CGAATGGCCA AGTACCAGGC AGCTGTGTCC AAACAAAGCA GCTCAACCAA
951 CTATACAAAT GAGCTGAAAG CCAGTGGTGG CGAAATCAA ATTCTATAAA
1001 TGGAGCAAAA GGAGAAATGT CCCCCAGGTC CTGAGGTCTG CATCACCCAT
1051 CAGGAAGGGG AAAAGATTTC TGCAAAATGAG AATAGCCTGG CAGTCCGTTC
1101 CACCCCTGCC GAAGATGACT CCCGTGACTC CCAGGTAAAG AGTGAGGTTC
1151 AACAGCCTGT CCATCCCAAG CCACTAAGTC CAGATTCCAG AGCCTCCAGT
1201 CTTTCTGAAA GTTCTCCTCC CAAAGCAATG AAGAAGTTTC AGGCACCTGC
1251 AAGAGAGACC TGCGTGGAAT GTCAGAAGAC AGTCTATCCA ATGGAGCGTC
1301 TCTTGCCCAA CCAGCAGGTG TTTACATCA GCTGCTCCG TTGCTCCTAT
1351 TGCAACAACA AACTCAGTCT AGGAACATAT GCATCTTTAC ATGGAAGAAT
1401 CTATTGTAAG CCTCACTTCA ATCAACTCTT TAAATCTAAG GGCAACTATG
1451 ATGAAGGCTT TGGGCACAGA CCACACAAGG ATCTATGGGC AAGCAAAAAT
1501 GAAAACGAAG AGATTTTGA GAGACCAGCC CAGCTTGCAA ATGCAAGGGA
1551 GACCCCTCAC AGCCAGGGG TAGAAGATGC CCCTATTGCT AAGGTGGGTG
1601 TCCTGGCTGC AAGTATGGAA GCCAAGGCCT CCTCTCAGCA GGAGAAGGAA
1651 GACAAGCCAG CTGAAACCAA GAAGCTGAGG ATCGCCTGGC CACCCCCCAC
1701 TGAACCTGGA AGTTCAGGAA GTGCCTTGA GGAAGGGATC AAAATGTCAA
1751 AGCCCAAAAT GCCTCCTGAA GACGAAATCA GCAAGCCCGA AGTTCCTGAG
1801 GATGTCGATC TAGATCTGAA GAAGCTAAGA CGATCTTCTT CACTGAAGGA
1851 AAGAAGCCGC CCATTCAGTG TAGCAGCTTC ATTTCAAAGC ACCTCTGTCA
1901 AGAGCCCAAA AACTGTGTCC CCACCTATCA GGAAGGGCTG GAGCATGTCA
1951 GAGCAGAGTG AAGAGTCTGT GGGTGGAAAG GTTGCAAGAA GGAACCAAGT
2001 GGAATGTCCT AAGGCTTCTA AGAAGAATGG GAATGTGGGA AAAACAACCT
2051 GGCAAAAACA AGAATCTAAA GGAGAGACAG GGAAGAGAAG TAAGGAAGGT
2101 CATAGTTTGG AGATGGAGAA TGAGAATCTT GTAGAAAATG GTGCAGACTC
2151 CGATGAAGAT GATAACAGCT TCCTCAAACA ACAATCTCCA CAAGAACCCA
2201 AGTCTCTGAA TTGGTCGAGT TTTGTAGACA ACACCTTTGC TGAAGAATTC
2251 ACTACTCAGA ATCAGAAATC CCAGGATGTG GAACCTCTGG AGGGAGAAGT
```



```

2301 GGTCAAAGAG CTCTCTGTGG AAGAACAGAT AAAGAGAAAT CGGTATTATG
2351 ATGAGGATGA GGATGAAGAG TGACAAATTG CAATGATGCT GGGCCTTAAA
2401 TTCATGTTAG TGTTAGCGAG CCACTGCCCT TTGTCAAAT GTGATGCACA
2451 TAAGCAGGTA TCCCAGCATG AAATGTAATT TACTTGGGAG TAACTTTGGA
2501 AAAGAAATCC TTCTTAAAT CAAAAACAAA AAAAAAAC AAAAAACA
2551 CATTCTAAAT ACTAGAGATA ACTTTACTTA AATTCTTCAT TTTAGCAGTG
2601 ATGATATGCG TAAGTGCTGT AAGGCTTGTA ACTGGGAAA TATTCCACCT
2651 GATAATAGCC CAGATTCTAC TGTATTCCCA AAAGGCAATA TTAAGGTAGA
2701 TAGATGATTA GTAGTATATT GTTACACACT ATTTTGGAA TAGAGAACAT
2751 ACAGAAGGAA TTTAGGGGCT TAAACATTAC GACTGAATGC ACTTTAGTAT
2801 AAAGGGCACA GTTGTATAT TTTTAAATGA ATACCAATTT AATTTTITAG
2851 TATTTACCTG TTAAGAGATT ATTTAGTCTT TAAATTTTT AGGTTAATTT
2901 TCTTGTGTG ATATATATGA GGAATTTACT ACTTTATGTC CTGCTCTCTA
2951 AACTACATCC TGAACCTGAC GTCCTGAGGT ATAATACAAC AGAGCACTTT
3001 TTGAGGCAAT TGA AAAACCA ACCTACACTC TTCGCTGCTT AGAGAGATCT
3051 GCTGTCTCCC AAATAAGCTT TTGTATCTGC CAGTGAATTT ACTGTACTCC
3101 AAATGATTGC TTTCTTTCT GGTGATATCT GTGCTTCTCA TAATTACTGA
3151 AAGCTGCAAT ATTTTAGTAA TACCTTCGGG ATCACTGTCC CCCATCTTCC
3201 GTGTTAGAGC AAAGTGAAGA GTTTAAAGGA GGAAGAAGAA AGAAGTGTCT
3251 TACACCACCT GAGCTCAGAC CTCTAAACCC TGTATTTCCC TTATGATGTC
3301 CCTTTTTTGA GACACTAATT TTTAAATACT TACTAGCTCT GAAATATATT
3351 GATTTTTTATC ACAGTATTCT CAGGGTGAAA TTAACCAAC TATAGGCCTT
3401 TTTCTTGGGA TGATTTTCTA GTCTTAAGGT TTGGGGACAT TATAAAGTCTG
3451 AGTACATTG TTGTACACAG TTGATATTCC AAATGTATG GATGGGAGGG
3501 AGAGGTGTCT TAAGCTGTAG GCTTTTCTTT GTACTGCATT TATAGAGATT
3551 TAGCTTTAAT ATTTTITAGA GATGTAAAC ATTCTGCTTT CTTAGTCTTA
3601 CCTAGTCTGA AACATTTTGA TTCAATAAAG ATTTTAATTA AAATTGAAA
3651 AAAAAAAAAA AAAA

```

BLAST Results

Entry HS512217 from database EMBL:

human STS SHGC-14654.

Length = 250

Minus Strand HSPs:

Score = 1202 (180.3 bits), Expect = 1.8e-46, P = 1.8e-46

Identities = 242/244 (99%)

Medline entries

95263566:

Three different rearrangements in a single intron truncate sterol regulatory element binding protein-2 and produce sterol-resistant phenotype in three cell lines. Role of introns in protein evolution.

93258417:

Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein.

Peptide information for frame 1

ORF from 94 bp to 2370 bp; peptide length: 759
Category: similarity to known protein

```

1 MESSPFNRQQ WTSLSLRVTA KELS LVNKNK SSAIVEIFSK YQKAAEETNM
51 EKKRSNTENL SQHFRKGTLT VLKKKWNPG LGAESHTDSL RNSSTEIRHR
101 ADHPPAEVTS HAASGAKADQ EEQIHPRSL RSPPEALVQG RYPHIKDGED
151 LKDHSTESKK MENCLGESRH EVEKSEISEN TDASGRIEY NVPLNRLKMM
201 FEKGEPTQTK ILRAQSRAS GRKISENSYS LDDLEIGPGQ LSSSTFDSEK
251 NESRRNLELP RLSETSIKDR MAKYQAAVSK QSSSTNYTNE LKASGGGEIKI
301 HKMEQKENVP PGPEVCITHQ EGEKISANEN SLAVRSTPAE DDSRDSQVKS
351 EVQQPVHPKP LSPDSRASSL SESSPPKAMK KFQAPARETC VECQKTVYPM
401 ERLLANQQVF HISCFRCSSY NNKLSLGYA SLHGRIYCKP HFNQLFKSKG
451 NYDEGFHHRP HKDLWASKNE NEEILERPAQ LANARETPHS PGVEDAPIAK
501 VGVLAASMEA KASSQKEKED KPAETKKLRI AWPFPTELGS SGSAALEEGIK
551 MSKPKWPPED EISKPEVPED VDLDLKLRR SSSLKERSRP FTVAASFQST
601 SVKSPKTVSP PIRKGWSMSE QSEESVGGRR AERKQVENAK ASKKNNGVKG
651 TTWQNKESKG ETGKRSKEGH SLEMENENLV ENGADSDDED NSFLKQQSPQ
701 EPKSLNWSFF VDNTFAEEFT TQNKQSQDVE LWEGEVVKEL SVEEQIKRNR

```


751 YYDEDEDEE

BLASTP hits

Entry CG22818_1 from database TREMBL:
 "SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. Cricetulus griseus (Chinese hamster)
 Length = 839
 Score = 1502 (528.7 bits), Expect = 3.9e-154, P = 3.9e-154
 Identities = 290/380 (76%), Positives = 322/380 (84%)

Entry S28507 from database PIR:
 transcription factor SF3 - common sunflower
 Length = 219
 Score = 212 (74.6 bits), Expect = 6.3e-18, Sum P(2) = 6.3e-18
 Identities = 36/82 (43%), Positives = 55/82 (67%)

Entry NTLIMDOM_1 from database TREMBL:
 "SF3"; product: "LIM-domain SF3 protein"; N.tabacum mRNA for LIM-domain protein Nicotiana tabacum (common tobacco)
 Length = 189
 Score = 216 (76.0 bits), Expect = 1.0e-16, P = 1.0e-16
 Identities = 42/94 (44%), Positives = 57/94 (60%)

Alert BLASTP hits for DKFZphut1_18i19, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18i19, frame 1

Report for DKFZphut1_18i19.1

[LENGTH] 759
 [MW] 85225.57
 [pI] 6.41
 [HOMOL] TREMBL:CG22818_1 gene: "SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. 1e-151
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR257w] 3e-05
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
 [BLOCKS] BL00478B
 [PIRKW] zinc finger 9e-16
 [PIRKW] DNA binding 9e-16
 [SUPFAM] LIM metal-binding repeat homology 9e-16
 [PROSITE] MYRISTYL 6
 [PROSITE] LIM_DOMAIN_1 1
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 28
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 15
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] LIM domain containing proteins
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 5.53 %

SEQ MESSPFNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL
 SEG
 lctl-
 SEQ SQHFRKGTLTVLKKKWENPGLGAESHTDSLRSSTEIRHRADHPPEVTSAAAGAKADQ
 SEG
 lctl-
 SEQ EEQIHPRSLRSPPEALVQGRYPHIKDGEDLKDSTESKKMENCLGESRHEVEKSEISEN
 SEG
 lctl-
 SEQ TDASGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRASGRKISENSYSLDDLEIGPGQ
 SEG


```

lctl- .....
SEQ  LSSSTFDSEKNESRRNLELPRLSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGGEIKI
SEG  .....
lctl- .....
SEQ  HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKSEVQQPVHPKP
SEG  .....x
lctl- .....
SEQ  LSPDSRASSLSESSPPKAMKFKQAPARETCVECQKTVYPMERLLANQQVFHISCFRCSYC
SEG  xxxxxxxxxxxxxxxx.....
lctl- .....ETTTTEETTTTCEEEETEEEEETTTTBTITT
SEQ  NNKLSLGTYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEEILERPAQ
SEG  .....
lctl- TCBCBTTBEEEEETEEEEETTTTTTTTTTCTTTTTTCTTT.....
SEQ  LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAWPPPTLGS
SEG  .....
lctl- .....
SEQ  SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPFTVAASFQST
SEG  .....xxxxxxxxxxxxxxxxxxxxx.....
lctl- .....
SEQ  SVKSPKTVSPPIRKGWSMSEQSEESVGRVAERKQVENAKASKKNGNVGKTTWQNKESKG
SEG  .....
lctl- .....
SEQ  ETGKRKSEGHSLMENENLVENGADSDDEDDNSFLKQOSPQEPKSLNWSSEVDNTFAEEFT
SEG  .....
lctl- .....
SEQ  TQNQKSQDVELWEGEVVKELSVEEQIKRNRYYDEDEDEE
SEG  .....xxxxxxx
lctl- .....

```

Prosites for DKFzphut1_18i19.1

PS00001	29->33	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00001	286->290	ASN_GLYCOSYLATION	PDOC00001
PS00001	706->710	ASN_GLYCOSYLATION	PDOC00001
PS00004	52->56	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	65->69	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	222->226	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	579->583	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	184->187	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	253->256	PKC_PHOSPHO_SITE	PDOC00005
PS00005	266->269	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	583->586	PKC_PHOSPHO_SITE	PDOC00005
PS00005	601->604	PKC_PHOSPHO_SITE	PDOC00005
PS00005	604->607	PKC_PHOSPHO_SITE	PDOC00005
PS00005	642->645	PKC_PHOSPHO_SITE	PDOC00005
PS00005	662->665	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS00006	294->298	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	326->330	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006

PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00006	514->518	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	658->662	CK2_PHOSPHO_SITE	PDOC00006
PS00006	686->690	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	709->713	CK2_PHOSPHO_SITE	PDOC00006
PS00006	714->718	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00007	223->230	TYR_PHOSPHO_SITE	PDOC00007
PS00007	222->230	TYR_PHOSPHO_SITE	PDOC00007
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	502->508	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	548->554	MYRISTYL	PDOC00008
PS00008	627->633	MYRISTYL	PDOC00008
PS00009	220->224	AMIDATION	PDOC00009
PS00009	662->666	AMIDATION	PDOC00009
PS00478	390->425	LIM_DOMAIN_1	PDOC00382

Pfam for DKFZphut1_18119.1

HMM_NAME	LIM domain containing proteins		
HMM	*CagCNrpIyDREivMRAMNKvWHpECFrCcdCqqPLtegdeFYErDGRI		
	C	C++++Y+ E++ A+ V+H++CFRC+ C+ L+ G+ + ++ GRI	
Query	390	CVECQKTVYPMERLL-ANQQVFHISCFRCSYCNNKLSLGT-YASLHGRI	436
HMM	YCKhDYrrFg*		
	YCK+++ ++F+		
Query	437	YCKPHFNQLEK	447

DKFZphut1_18i4

group: uterus derived

DKFZphut1_18i4 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

weak similarity to C.elegans D2085.2

complete cDNA, complete cds, few EST hits

Sequenced by AGOWA

Locus: /map="7q31"

Insert length: 1568 bp

Poly A stretch at pos. 1551, polyadenylation signal at pos. 1523

```
1  GCCGAGCGGA GAGGGTAGAG ACGGGGTTTC ACCGTGTTAG CCAAGATGGT
51  CTCGATCTCC TGACCTCGTG ATCCGCCCCG CTCGGCCTCC CAAAGTGCTG
101 GGATTACAGG CGTGAGCCAC TGCGCCCGGC CTGTTGTACA GTTATTAAAG
151 TTATCATTTA ACATGGAAGA AGATGAGTTC ATTGGAGAAA AAACATTCCA
201 ACGTTATTGT GCAGAATTCA TTAACATTTC ACAACAGATA GGTGATAGTT
251 GGGAAATGGAG ACCATCAAAG GACTGTCTG ATGGCTACAT GTGCAAAATA
301 CACTTTCAAA TTAAGAATGG GTCTGTGATG TCACATCTAG GAGCATCTAC
351 CCATGGACAG ACATGTCTTC CCATGGAGGA GGCTTTCGAG CTACCCCTGG
401 ATGATTGTGA AGTGATTGAA ACTGCAGCAG CGTCCGAAGT GATTAAATAT
451 GAGTATCATG TCTTATATTC CTGTAGCTAC CAAGTGCCTG TACTTTACTT
501 TAGGGCAAGC TTTTATAGAT GGAGACCTTT AACTCTGAAG GACATATGGG
551 AAGGAGTTCA TGAGTGCTAT AAGATGCGAC TGCTACAGGG ACCATGGGAC
601 ACTATTACGC AACAGGAACA TCCAATACTT GGGCAACCCT TTTTGTACT
651 TCATCCCTGC AAGACGAATG AATTCATGAC TCCTGTATTA AAGAATTCTC
701 AGAAAAATCA TAAGATGTC AACTATATCA CATCATGGCT GAGCATTGTA
751 GGGCCAGTTG TTGGGCTGAA TCTACCTCTG AGTTATGCCA AAGCAACGTC
801 TCAGGATGAA CGAAATGTCC CTTAACAAGA TTCTTCTATT GAGTTTAGGA
851 ATTGCGGCAC GAAGAATGCC AAGAGTTTAC CTGGCCAGCC CTGGCTTTAA
901 TAGGACTGAT ACCATGGAAT ATTTATCTC ACCAAGATGT GACATGGATT
951 ATTTTCCCT TGGACACAAA TGTCTACAGC AACTGATGTT TGATAGGCTG
1001 AATGTTTAGA AGAAACACTT CAAAGGGATA CATCATGGCC AGGCATGGTG
1051 GCTCACACCT GTAATCCAAG CACTTTGGGA GGCCAAGGTG GGAGCATCAC
1101 TTGATCCTGG GAGTTCGAGA CCAGCCTGGG CAACATGGTG AAACCTGTC
1151 GGTACAAAAA AATACAAAAA TTTGCCTGTT TATGGTGGTG TGTTCCTGTA
1201 GTCCCAGCTC CCCAGGAGGC TGAGGTGGGA GGTGGGCTTT AACCCAGGAG
1251 GCAGAGGTTG CAGTGAGCTG AGACTGTGCC ACTGCAGTCC AGCCTGGGTG
1301 ACAGAGCCAG ACACTGTCTC GGGAAAAAAA AAAAAAAA AAAGACACAT
1351 CACTATAAAT AGCAAAAAA CAAATCTAAC TTATTAATAC TAGGAATACC
1401 AACATTATTA GGGCACTTGC AGGTTATCT TTTCTAGGCC AAGTACTTCA
1451 CTTCCATTG TCTGACATGG AGATTGAGGG AGAAATGTAT TTGTGTGTTT
1501 ATTTTAATGT AAGATATATA AAAATTAAAT TACTGGATT ACCTGTCCCT
1551 GAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 163 bp to 822 bp; peptide length: 220
Category: similarity to unknown protein

BLASTP hits

Alert BLASTP hits for DKFZphut1 18i4, frame 1

Pedant information for DKFZphutet_18i4, frame 1

Report for DKFZphutel 18i4.1

[BLOCKS]	BL00221E	
[PROSITE]	MYRISTYL	2
[PROSITE]	CK2_PHOSPHO_SITE	4
[PROSITE]	PKC_PHOSPHO_SITE	2
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha Beta	

[illegible]

Prosites for DKFZphutel1_18i4.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00005	124->127	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	131->137	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_18i4.1)

DKFZphut1_1811

group: nucleic acid management

DKFZphut1_1811 encodes a novel 184 amino acid protein with similarity to *S. cerevisiae* putative ribosomal protein YHR148w.

The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome subunit.

The new protein can find application in modulation of ribosome assembly, structure and function.

strong similarity to *S.cerevisiae* YHR148w

complete cDNA, complete cds, EST hits,
potential start at Bp 45 matches kozak consensus ANNatgG
gene disruption of YHR148w is lethal!

Sequenced by AGOWA

Locus: unknown

Insert length: 1076 bp

Poly A stretch at pos. 1035, polyadenylation signal at pos. 1006

```
1 GCGCGCTCTC AGCTTCGGGT CCTGCGGCTG CGGCTGCCGC CATCATGGTG
51 CGGAAGCTTA AGTTCCACGA GCAGAAGCTG CTGAAGCAGG TGGACTTCCT
101 GAAGCTGGGAG GTCACCGACC ACAACCTGCA CGAGCTGCCG GTGCTGCCGC
151 GTTACCGGCT GCAGCGGCGG GAGGACTACA CGCGCTACAA CCAGCTGAGC
201 CGTGCCGTGC GTGAGCTGGC GCGCGCGCTG CGCGACCTGC CCGAACGCGA
251 CCAGTTCCGC GTGCGCGCTT CGGCCGCGCT GCTGGACAAG CTGTATGCTC
301 TCGGCTTGGT GCCCACGCGC GGTTCTGCTG AGCTCTGCGA CTTCGTCACG
351 GCCTCGTCCT TCTGCCGCGC CCGCCTCCCC ACCGTGCTCC TCAAGCTGCG
401 CATGCCGCGC CACCTTCAGG CTGCCGTGGC CTTTGTGGAG CAAGGGCACG
451 TACGCGTGGG CCTGACGTG GTTACCGACC CCGCCTTCCT TGTACCGCGC
501 AGCATGGAGG ACTTGTGTC TGGGTGGGAC TCGTCCAAGA TCAAGCGGCA
551 CGTCTAGAG TACAATGAGG AGCGCGATGA CTTGATCTG GAAGCCTAGC
601 GGATCTCCCA CTTTGCATGC CTGTCTTTTA CAGATGGGAA AACTGAGGCC
651 TGATGCTGGA GATTCTATGA GGGTGCTCTC CTCGAGGGTA TCAGACGGTC
701 GTAGGTTCTT AAGAATTTGA TTCATCAGTG GCAGGCCATG CATAGAGCCA
751 CCGGAGGTGC GTCCTTGTIT TCCAGGAAAT GTTCTTAGAA CTTGGACTAC
801 TGATTATTAA TTGACTGTGC CTTGGGAAAC AGTGGGAAGT AACTTGGTGC
851 AGCACTGGGG TATTGTTGGA CTGGTTCAAT TCGTTTAACT CGAATTCTTG
901 CTCCTGGCGG TGGTTAAGCT GTGTACAGAT GATGGAGAGT TTGGCCTCAA
951 GTTTTATAAA ACTGAGCGAG ACTAGTGTTT AGGATCTCCT CCCTTGTTTA
1001 AATGTCAATA AATGCCCAA CTGCTTTGTA AGCTCAAAAA AAAAAAAAAA
1051 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 45 bp to 596 bp; peptide length: 184
Category: strong similarity to known protein

```
1 MVRKLFHEQ KLLKQVDFLN WEVTDHNLHE LRLVRRYRLO RREDYTRYNQ
51 LSRVRELAR RLRDLPERDQ FRVRASAALL DKLYALGLVP TRGSLELCDF
101 VTASSFCRRR LPTVLLKLRM AQHLQAAVAF VEQGHVRVGP DUVTDPAFLV
151 TRSMEDEVTV VDSSKIKRHV LEYNEERDDF DLEA
```

BLASTP hits

Report for DKFZphut1_1811.3

```

[LENGTH]      184
[MW]           21850.21
[pI]           9.54
[HOMOL]        PIR:S33911 probable ribosomal protein YHR148w - yeast (Saccharomyces
cerevisiae) 4e-47
[FUNCAT]       05.01 ribosomal proteins [S. cerevisiae, YHR148w] 2e-48
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YPL081w] 5e-07
[FUNCAT]       j m rna translation and ribosome biogenesis [M. jannaschii, MJ0190] 8e-05
[BLOCKS]       BL00632
[PIRKW]        cytosol 1e-07
[PIRKW]        ribosome 1e-07
[PIRKW]        protein biosynthesis 1e-07
[SUPFAM]       rat ribosomal protein S9 1e-07
[PROSITE]      MYRISTYL 1
[PROSITE]      CK2_PHOSPHO_SITE 2
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 1
[PFAM]         Ribosomal protein S4
[KW]           All Alpha
[KW]           LOW COMPLEXITY 6.52 %

```

```
SEQ      MVRKLFKEQKLLKVDFLNWEVTDHNLHELRLVRRYRLQRREDYTRYNQLSRAVRELAR  
SEG      .....xxxxxxxxxxxxx.....  
PRD      cccchhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
  
SEQ      RLRDLPERDQFRVRASAAALLDKLYALGLVPTRGSLELCDFVTASSFCRRRLPTVLLKLRM  
SEG      .....  
PRD      hhhhhccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
  
SEQ      AQHLQAAVAFVEQGHRVGPDPVDTDPFLVTRSMEDFVTWVDSSKIKRHVLEYNEERDDF  
SEG      .....  
PRD      hhhhhhhhhhhhhhhccccceeeccceeeeeccceeeeeccchhhhhhhhhcccccc  
  
SEQ      DLEA  
SEG      ....  
PRD      CCCC
```

Prosites for DKFZphute1_1811.3

PS000005	163->166	PKC_PHOSPHO_SITE	PDOC000005
PS000006	153->157	CK2_PHOSPHO_SITE	PDOC000006
PS000006	159->163	CK2_PHOSPHO_SITE	PDOC000006
PS000007	41->49	TYR_PHOSPHO_SITE	PDOC000007
PS000008	87->93	MYRISTYL	PDOC000008

Pfam for DKFZphute1 1811.3

HMM_NAME	Ribosomal protein S4		
HMM		*MSR.YRGP ^W KW ^K II ^R RP ^G ELP ^W LTnK....tkl ^m r ^k yC..l ^R PG ^Q H ^G W ^R	
Query	1	M ⁺ R ++ +++K++++++L W +++++R Y R+++ ++ MVRKLK ^F HEQ ^K LL ^K Q ^V DFLN ^E W ^E TD ^H NL ^H EL ^R VL ^R RY ^R L ^Q RR ^E DY ^T RYN	49
HMM		qRktLsKI ^R Rm ^S QY ^r IRLQE ^K K ^L RF ^M YGN ^I TER ^Q L ^R RY ^V riaEd ^K R ^K L ^D	
Query	50	Q + +R +++ + L+E + +R +++++L++++ + +L QLSR--AVRELAR ^R LRDL ^R PERD ^Q FRV ^R ASA ^A LLDK ^L YAL ^G LV ^P -TRGS ^L	96
HMM		YsTGEnLM ^Q ILE ^M RL ^D NI ^V FR ^M GM ^A PTI ^H HA ^R QLIN ^H R ^H IR ^V Nd ^R IV ^N IP	
Query	97	++ +++++RL++++ ++ MA ++A+ +++++H+RV++ +V++P LCDFVTASS ^F CR ^R RL ^P TV ^L L ^K LR ^M AQ ^H LQ ^A AV ^F VE ^Q GH ^V RV ^G PD ^V VT ^D P	146
HMM		SYiCR ^P NDi ^I SI ^R dkqrMQ ^S Hi ^k W ⁿ ieSPeg ^r mRPN ^H LE ^r N ⁿ kk ^y eG ^t IN	


```
Query      147  ++++++ +      +++++W++ S+      ++R+ + Y+ +  
AFLVTRS---M-----EDFVTWVDSSK-----IKRHVLEYNEERD 178  
HMM      rIIEReWiplkINeLLVVEY*  
Query      179  +++ +  
DFDLE----- 183
```


DKFZphutel_19f19

group: transmembrane protein

DKFZphutel_19f19 encodes a novel 204 amino acid protein with similarity to murine p24 protein.

Murine p24 is expressed only in brain where it is localized exclusively in neurons. It seems to be a neuron-specific membrane protein localised in intracellular organelles of highly differentiated neural cells and may play a role in the neural organelle transport system. As p24, the novel protein contains 2 transmembrane regions, but it contains not the sequence homologous to the microtubule-binding domain of microtubule-associated proteins present in p24.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to mouse P24 protein ;

membrane regions: 2

Summary DKFZphutel_19f19 encodes a novel 204 amino acid protein, with similarity to mouse P24 protein.

similarity to mouse P24 protein

complete cDNA, complete cds, EST hits,
2 TM-domains

Sequenced by AGOWA

Locus: /map=14.8 cR from top of Chr20 linkage group

Insert length: 2042 bp

Poly A stretch at pos. 1958, polyadenylation signal at pos. 1940

```

1 GCAGGCAGAG AGATGAGGAA ACTGAGACCC AGAAAGGTGG AAGCACTTGT
51 CTAAGGTCAC GCCTCCAGGA AGCAGTGTGT CCACGACTCC AGTCCAAGTG
101 GTCAGGCTCC AGAGCCACACA GTCCCAGGGG TCCATGATGC CGAGCTGCAA
151 TCGTTCTCTGC AGCTGCAGCC GCGGCCCCAG CGTGGAGGAT GGCAGTGGT
201 ATGGGGTCCG CTCTACCTG CACCTCTTCT ATGAGGACTG TGCAGGCACT
251 GCTCTCAGCG ACGACCTGA GGGACCTCCG GTCCTGTGCC CCCGCCGGCC
301 CTGGCCCTCA CTGTGTTGGA AGATCAGCCT GTCCTCGGGG ACCCTGCTTC
351 TGCTGCTGGG TGTGGCGGCT CTGACCACTG GCTATGCAGT GCGCCCAAG
401 CTGGAGGGCA TCGGTGAGGG TGAGTTCCTG GTGTTGGATC AGCGGGCAGC
451 CGACTACAAC CAGGCCCTGG GCACCTGTCT CCTGGCAGGC ACAGCGCTCT
501 GTGTGGCAGC TGGAGTCTCT CTCGCCATCT GCCTCTTCTG GGCCATGATA
551 GGCTGGCTGA GCCAGGACAC CAAGGCAGAG CCCTTGGACC CCGAAGCCGA
601 CAGCCACGTG GAGGTCTTCG GGGATGAGCC AGAGCAGCAG TTGTACCCCA
651 TTTTCCGCAA TGCCAGTGGC CAGTCATGGT TCTCGCCACC CGCCAGCCCC
701 TTTGGGCAAT CTTCTGTGCA GACTATCCAG CCCAAGAGGG ACTCCTGAGC
751 TGCCACATG GCCTAAGATG TGGGTCTCTG ATCCTTCCCC CTCTCACCAC
801 TAACCCCTCT TCAGTGTTC CCCAACTTCT CCCTTAGAG CCCAACTCCA
851 GGTCAAATCT GGAGCTCAA TCCCAGTGCT CCCTCCCCAG GAGTGGGGCC
901 CCAACTCTTC CAAGATACCA GCATTCTCTA AGTCCTCCCA AAACCTTCTA
951 CCCACACCCT CTTCCCAAGG CCCTCAGGGG CAGAAAACAT CTCCTTCAAC
1001 CCGTCCCCAC TCCTTCTCTT GCATGACCTT GGGCAAAACC TTGCCCTTTC
1051 AAGCCATCAG CTCCTGCCCT TCTGCCATGA GGGCTTTGGA TCAGATTCTT
1101 CTTCTCGCCA GGATGAGGAC ACGCACTGCC CTCCATAGAC ACAGATGAAG
1151 GGGTGGGGGT CATTAGCTCT GAATGGGTCC CAGATGTCTA CTTGGCCTTT
1201 CCCTGCAGGA TGAGTGAAGA CGTTTGCTCT TCACAGTGTG TCTTCTACCT
1251 GCATTTTGGC ATCAGAGCCC CCCAGCCCAC CCACCACAGG CAATTACTAG
1301 CCCTAGTTGA TAGGTGAGGT GGGTGAAGAA GGCTGGAGGT GACATGTCCG
1351 AGGTCACACA ACAAGCAGC ATGCAGGAAC TAGAAACACA TCTTCAGCCT
1401 CCTCCTGGGC CAGCTCTTGT GCTACAGGTG GGGCGGAGCC AGCCCTCAC
1451 CTTCTGGGTT CCCTGAGGGT CCTCAGGGTG GAGGACAGGT TTGGCCCGAG
1501 AAGACTAGCC AGAGGCCTGA TGGTCCCAGG TGGCTCTGGA TATACTTTGG
1551 ATATGGATTT AAATGGTCTC TAAGAGCCGG GGGTAGGGGG CAGGAAAAGT
1601 GGGTGTGCTT TGCCCTCAA AGTCCACCTA CCTAGAAACC AAGCCCACGG
1651 TCTTGGCCGT GACCCGTGTA ATAAATGGGC TCTCTCAGAG GCGCCAGCCC
1701 CTCCTCTCCC AGCCGGAGGC GTCATCTCTC TTCTGTACCA CTAGAGGGAG
1751 CTTCTGATGA GCTGGAGAGC AGCGCTCAAG GCTCTCGCCC CTCCCTCCCC
1801 TAACCCCTTA CTTTCACTCT CACCAGCCTG AAGGGCCTCC TAGGGGATCC
1851 TCAGGCGGGC CCCACCAGGG CACACCCTAC TGTCTTGTG CCTCAGCCCC
1901 CCTCCTCATC CTGCAACCCCT TCCATCCAC CTTCCTTTTC AATAAACAGC
1951 TGGGATGGAA AAAAAAAAAA AGAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2001 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA

```


BLAST Results

Entry HS417348 from database EMBL:
human STS WI-14697.
Length = 290
Minus Strand HSPs:
Score = 1254 (188.2 bits), Expect = 3.0e-50, P = 3.0e-50
Identities = 262/273 (95%)

Medline entries

97334404:
A newly identified membrane protein localized exclusively in
intracellular organelles of neurons.

Peptide information for frame 2

ORF from 134 bp to 745 bp; peptide length: 204
Category: similarity to known protein

1 MMPSCNRSCS CSRGPSVEDG KYGVRSYLH LFYEDCAGTA LSDDPEGPPV
51 LCPRRPWPSL CWKISLSSGT LLLLLGVAAL TTGYAVPPKL EGIGEGEFLV
101 LDQRAADYNG ALGTCRLAGT ALCVAAGVLL AICLFWAMIG WLSQDTKAEP
151 LDPEADSHVE VFGDEPEQQL SPIFRNASGQ SWFSPPASPF GQSSVQTIQP
201 KRDS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphute1_19f19, frame 2

TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
complete cds., N = 1, Score = 295, P = 3.8e-26

>TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
complete cds.
Length = 196

HSPs:

Score = 295 (44.3 bits), Expect = 3.8e-26, P = 3.8e-26
Identities = 58/139 (41%), Positives = 81/139 (58%)

Query: 2 MPSCNRSCSCSRGPSVEDGKW---YGVRSYLHLYFYEDCAGTALSDDPEGPPVLCPRRPWP 58
M SC+ +C R + +G + YGVRSYLH FYEDC + + + P R W
Sbjct: 1 MTSCSNTCGSRRAQADTEGGYQQRYSYLYHQFYEDCTASIWEYEDDFQIQSPNR-WS 59

Query: 59 SLCWKISLSSGTLLLLLGVAALTTGYAVPPKLEGIGEGEFLVLDQRAADYNQALGTCRLA 118
S+ WK+ L SGT+ ++LG+ L G+ VPPK+E GE +F+V+D A YN AL TC+LA
Sbjct: 60 SVFWKVGILISGTVFVILGLTVLAVGFLVPPKIEAFGEADFMVVDTHAVKYNALDTCKLA 119

Query: 119 GTALCVAAGVLLAICLFWAM 138
G L G +A CL ++
Sbjct: 120 GAVLFCIGGTSMAGCLLMSV 139

Pedant information for DKFzphute1_19f19, frame 2

Report for DKFzphute1_19f19.2

[LENGTH] 204
[MW] 21983.07
[pI] 4.69
[HOMOL] TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein, complete
cds. 7e-19
[PROSITE] MYRISTYL 4

DKF2phutel_19g19

group: uterus derived

DKF2phutel_19g19 encodes a novel 400 amino acid protein, with strong but partial similarity to a bovine elastin-related protein expressed in fetal calf ligamentum nuchae.

The novel protein contains 2 RGD cell attachment sites.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to bovine elastin fragment

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: map=54.9 cR from top of Chr3 linkage group

Insert length: 3244 bp

Poly A stretch at pos. 3227, polyadenylation signal at pos. 3216

```
1  GTAAGTGCAG TAAGTCCCGC TTGGCCCTGG AGTCCACGCG GATTTTCGAA
51  GCTGGGGGCTG GCAAGAGGCC GCTGGACACC ACGCTCCAGT CGTCAGCCCA
101 CTTCTTAGCTG GAACAGCGCG AGGCGGCGCG AGCGAGCCGG GTCCCACCAT
151 GGCCGCGAAT TATTCCAGTA CCAGTACCCG GAGAGAACAT GTCAAAGTTA
201 AAACCAGCTC CCAGCCAGGC TTCTGGGAC GGCTGAGCGA GACCTCGGGT
251 GGGATGTTTG TGGGGCTCAT GGCCTTCCTG CTCTCCTTCT ACCTAATTTT
301 CACCAATGAG GGCCGCGCAT TGAAGACGGC AACCTCATTG GCTGAGGGGC
351 TCTCGCTTGT GGTGTCTCCT GACAGCATCC ACAGTGTGGC TCCGGAGAAT
401 GAAGGAAGGC TGGTGCACAT CATTGGCGCC TTACGGACAT CCAAGCTTTT
451 GTCTGATCCA AACTATGGGG TCCATCTTCC GGCTGTGAAA CTGCGGAGGC
501 ACGTGGAGAT GTACCAATGG GTAGAACTG AGGAGTCCAG GGAGTACACC
551 GAGGATGGGC AGGTGAAGAA GGAGACGAGG TATTCTACA ACACTGAATG
601 GAGGTCAAGG ATCATCAACA GCACAAACTT CGACCGAGAG ATTGGCCACA
651 ATAACCCGAG TGCCATGGCA GTGGAGTCAT TCACGGCAAC AGCCCCCTTT
701 GTCCAAATG GCAGGTTTTT CCTCTCGTCA GGCCTCATCG ACAAAGTCGA
751 CAATCTCAAG TCCCTGAGCC TATCCAAGCT GGAGGACCCCT CATGTGGACA
801 TCATTTCGCC TGGAGACTTT TTCTACCACA GCGAAAATCC CAAGTATCCA
851 GAGGTGGGAG ACTTGCCTGT CTCCTTTTCC TATGCTGGAC TGAGCGGCGA
901 TGACCCCTGAC CTGGGCCCAG CTCACGTGGT CACTGTGATT GCCCGGCAGC
951 GGGGTGACCA GCTAGTCCCA TTCTCCACCA AGTCTGGGGA TACCTTACTG
1001 CTCTGACACC ACGGGGACTT CTCAGCAGAG GAGGTGTTTC ATAGAGAACT
1051 AAGGAGCAAC TCCATGAAGA CCTGGGGCCT GCGGGCAGCT GGCTGGATGG
1101 CCATGTTTAT GGGCCTCAAC CTTATGACAC GGATCCTCTA CACCTTGGTG
1151 GACTGGTTTC CTGTTTCCG AGACCTGGTC AACATTGGCC TGAAGACCTT
1201 TGCCTTCTGT GTGGCCACCT CGCTGACCCT GCTGACCCTG GCGGCTGGCT
1251 GGCTCTTCTA CCGACCCCTG TGGGCCCTCC TCATTGCCGG CCTGGCCCTT
1301 GTGCCCATCC TTGTTGCTCG GACACGGGTG CCAGCCAAAA AGTTGGAGTG
1351 AAAAGACCTT GGCACCCGCG CGACACCTGC GTGAGCCCTA GGATCCAGGT
1401 CCTCTCTCAC CTCTGACCCA GCTCCATGCC AGAGCAGGAG CCCCAGTCAA
1451 TTTTGGAGTC TGCACCCCTT CTCCTCTTCA GGGGCCAGAC TTGGCAGCAT
1501 GTGCACCAGG TTGGTGTTC CAGCTCATG TCTTCCCCAC ATCTCTTCTT
1551 GCCAGTAAGC AGCTTGTGTG GGCAGCAGCA GCCATGAATG GCAAGCTGAC
1601 AGCTTCTCCT GCTGTTTCCT TCCTCTCTTG GACTGAGTGG GTACGGCCAG
1651 CCACTCAGCC CATTGGCAGC TGACACGCA GACACGCTCT ACGGAGGCCCT
1701 GCTGATAAAG GGCTCAGCCT TGCCGTGTGC TGCTTCTCAT CACTGCACAC
1751 AAGTGCCATG CTTTGCCACC ACCACCAAGC ACATCTGTGA TCCTGAAGGG
1801 CGGCCGTTAG TCATTACTGC TGAGTCTTGG GTCACCAGCA GACACACTGG
1851 GCATGGACCC CTCAAAGCAG GCACACCCAA AACACAAGTC TGTGGCTAGA
1901 ACCTGATGTG GTGTTTAAAA GAGAAGAAAC ACTGAAGATG TCCTGAGGAG
1951 AAAAGCTGGA CATATACTGG GCTTCACACT TATCTTATGG CTGGCCAGAA
2001 TCTTTGTAGT GTGTGGGATC TCTGAAGGCC CTATTTAAGT TTTTCTTCGT
2051 TACTTTGCTG CTTATGTGT ACTTTCCTAC CCCAAGAGGA AGTTTTCTGA
2101 AATAAGATTT AAAAACAAAA CAAAAAAAC ACTTAATATT TCAGACTGTT
2151 ACAGGAAACA CCGTTTAGTC TGTCAGTTGA ATTCAGAGCA CTGAAAGGTG
2201 TTAATTTGGG GTATGTGGTT TGATTGATAA AAAGTTACCT CTCAGTATTT
2251 TGTGTCACTG AGAAGCTTTA CAATGGATGC TTTTGAACA AGTATCAGCA
2301 AAAGGATTTG TTTTCACTCT GGGAGGAGAG GGTGGAGAAA GCACCTTGCTT
2351 TCATCTCTCT GCATCGGAAA CTCCCTATG CACTTGAAGA TGGTTTAAAA
2401 GATTAAAGAA ACGATTAAGA GAAAGGTTG GAAGCTTTAT ACTAAATGGG
2451 CTCCTTCAAT GTGACGCCCC GTCAACCACA ATCAAGAACT GAGGCCTGAG
2501 GCTGGTTGTA CAATGCCAC GCCTGCCTGG CTGCTTTCAC CTGGGAGTGC
2551 TTTGATGTG GGCACCTGGG CTTCCTAGGG CTGCTTCTGA GTGGTTCTTT
2601 CACGTGTTGT GTCCATAGCT TTAGTCTTCC TAAATAAGAT CCACCCACAC
```



```

2651 CTAAGTCACA GAATTTCTAA GTTCCCAAC TACTCTCACA CCCTTTTAAA
2701 GATAAAGTAT GTTGTAACCA GGATGTCTTA AATGATTCTT TGTGTACCTT
2751 TTCTGTGATA TTCAGAAACC GTTTGTGCC TGCTGGGAGT AATTCCTTTA
2801 GCAATTAAGT ATTTGGTAGC TGAATAAGGG GTCAGAACTT CTGAAACCAG
2851 AGATCTGTAA TCATCTCTAT TGGCCTGGGG TGCCTGTGCT ATAAATGAGT
2901 TTCTTCACAT GAAAAACACA GCCAGCCCAA GATGACTTAT CTGGGTTTAG
2951 GATTCAATAG TATTCATAA CTGCTTATTA CATGAGCAAT TTCATCAAA
3001 CTCCAAACTC TTAAGGATG CTTTCGAAA ACACGCTGTA TACCTAGATG
3051 ATGACTAAAT GCAAAATCCT TGGGCTTTGG TTTTCTTCTA GTAAGGATTT
3101 TAAATAACTG CCGACTTCAA AAGTGTCTT AAAACGAAAG ATAATGTTAA
3151 GAAAAATTG AAAGCTTTGG AAAACCAAT TTGTAATATC ATTGTATTTT
3201 TTATTAAG TTTTGAATA AATTTCTAAA AAAAAAAAAA AAAA

```

BLAST Results

Entry HS545355 from database EMBL:
human STS WI-14815.
Length = 436
Minus Strand HSPs:
Score = 2040 (306.1 bits), Expect = 6.2e-86, P = 6.2e-86
Identities = 420/426 (98%)

Entry HS932147 from database EMBL:
human STS WI-8531.
Length = 341
Minus Strand HSPs:
Score = 1705 (255.8 bits), Expect = 4.7e-70, P = 4.7e-70
Identities = 341/341 (100%)

Medline entries

86051793:
Bovine elastin cDNA clones: evidence for the occurrence of a
new elastin-related protein in fetal calf ligamentum nuchae.

Peptide information for frame 2

ORF from 149 bp to 1348 bp; peptide length: 400
Category: similarity to known protein

```

1 MAANYSSSTST RREHVVKVTS SQPGFLERLS ETSGGMFVGL MAFLLSFYLI
51 FTNEGRALKT ATSLAEGLSL VVSPDSIHSV APENEGRLVH IIGALRTSKL
101 LSDPNYGVHL PAVKLRRHVE MYQWVETES REYTEDGQVK KETRYSYNTE
151 WRSEIINSKN FDREIGHNNP SAMAVESFTA TAPFVQIGRF FLSSGLIDKV
201 DNFKSLSLSK LEDPHVDIIR RGDFFYHSEN PKYPEVGDLR VSFSYAGLSG
251 DDPDLGPAHV VTVIARQRGD QLVPFSTKSG DTLILLHHGD FSAEEVFHRE
301 LRSNSMKTWG LRAAGWMAMF MGLNLMTRIL YTLVDWFPVF RDLVNIGLKA
351 FAFCVATSLT LLTVAAGWLF YRPLWALLIA GLALVPILVA RTRVPAKKLE

```

BLASTP hits

Entry I45887 from database PIR:
elastin - bovine (fragment)
Length = 40
Score = 131 (46.1 bits), Expect = 4.9e-08, P = 4.9e-08
Identities = 31/41 (75%), Positives = 34/41 (82%)

Alert BLASTP hits for DKFZphutel_19g19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphutel_19g19, frame 2

Report for DKFZphutel_19g19.2

[LENGTH] 400


```

[MW]          44831.53
[pI]          7.23
[HOMOL]       PIR:I45887 elastin - bovine (fragment) 1e-06
[PROSITE]     RGD      2
[PROSITE]     MYRISTYL  3
[PROSITE]     CAMP_PHOSPHO_SITE  1
[PROSITE]     CK2_PHOSPHO_SITE   6
[PROSITE]     TYR_PHOSPHO_SITE   2
[PROSITE]     PKC_PHOSPHO_SITE   5
[PROSITE]     ASN_GLYCOSYLATION  1
[KW]          TRANSMEMBRANE 4

SEQ  MAANYSSSTSTRREHVVKVKTSSQPGFLERLSETSGGMFVGLMAFLLSFYLIFTNEGRALKT
PRD  ccccecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  ATSLAEGLSLVVSPDSIHSVAPENEGRLVHIIGALRTSKLLSDPNYGVHLPVVKLRHVE
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  MYQWVETESREYTEDGQVKKETRYSYNTEWRSEIINSKNFDRIGHNNPVSAMAVESFTA
PRD  hheeehhhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....M

SEQ  TAPFVQIGRFFLSSGLIDKVDNFKSLSLSKLEDPHVDIIRRGDFFYHSENPKYPEVGDLR
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ  VSFSYAGLSGDDPDLGPAHVTVIARQRGDQLVPFSTKSGDTLLLLHHGDFSAAEVFHRE
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  LRSNSMKTWGLRAAGWMAMFMGLNLMTRILYTLVDWFPVFRDLVNIGLKAFACVATSLT
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMM

SEQ  LLTVAAGWLFYRPLWALLIAGLALVPILVARTRVPKLE
PRD  hhhhccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

Prosites for DKFZphut1_19g19.2

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	332->336	CK2_PHOSPHO_SITE	PDOC00006
PS00007	220->227	TYR_PHOSPHO_SITE	PDOC00007
PS00007	99->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00016	221->224	RGD	PDOC00016
PS00016	268->271	RGD	PDOC00016

(No Pfam data available for DKFZphut1_19g19.2)

DKFZphutel_19g22

group: cell structure and motility

DKFZphutel_19g22 encodes a novel 390 amino acid protein with very strong similarity to tuftelin/enamelin.

Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix.

The new protein can find application in modulation of tissue-calcification, especially the uterus.

complete cDNA, complete cds start at Bp 51, EST hits in 3' UTR,
human homolog of mouse tuftelin
tuftelin is described as a matrix protein of teeth but it seems also
to be pressend in the uterus matrix

Sequenced by AGOWA

Locus: unknown

Insert length: 3110 bp

Poly A stretch at pos. 3093, polyadenylation signal at pos. 3071

```
1 GCAGACAGCG GGGTGGACAA GTGGCGTGTG TGCTGCGACC CCGAGGGAAG
51 ATGAACGGGA CGCGGAAC TGTTACCTGT GTGGACGTGC ACCCAGAGGA
101 CCAGGCGGCG GGCAGCGTGG ACATTCTCAG GCTGACTCTC CAGGGTGAAC
151 TGACAGGAGA TGAAC TTGAA CACATAGCCC AGAAGGCGGG CAGGAAGACC
201 TATGCCATGG TGTCCAGCCA CTCAGCTGGT CATCTCTGG CTTCAGAACT
251 GGTGGAGTCC CATGATGGAC ATGAGGAGAT CATTAAAGGT TACTTGAAGG
301 GGAGGTCTGG AGACAAGATG ATTACAGAGA AGAATATTA CCAGCTGAAG
351 AGTGAGGTCC AGTACATCCA GGAGGCCAGG AACTGCCTAC AGAAGCTCCG
401 GGAGGATATA AGTAGCAAGC TTGACAGGAA CCTAGGAGAT TCTCTCCATC
451 GACAGGAGAT ACAGGTGGTG CTAGAAAAGC CAAATGGCTT TAGTCAGAGT
501 CCCACAGCCC TGTACAGCAG CCCACCTGAG GTGGACACCT GTATAAATGA
551 GGATGTTTGG AGCTTGAGGA AGACGGTGCA GGACTTGCTG GCCAAGCTTC
601 AGGAGGCCAA GCGCAACAC CAGTCAGACT GTGTGGCTTT TGAGGTGACA
651 CTCAGCCGGT ACCAGAGGGA AGCAGAACAA AGTAATGTGG CCCTTCAGAG
701 AGAGGAGGAG AGAGTGGAGC AGAAGAGGCG AGAAGTCGGA GAGCTGCAGA
751 GGCCTTTGCT AGGGATGGAG ACGGAGCATC AGGCCTTACT GGCGAAAGTG
801 AGGGAAGGGG AGGTGGCCCT AGAGGAACTT CGGAGCAACA ATGTGACTGT
851 CCAAGCAGAA CGAGAAAAGG CTGCTACCTT GGAAGAGGAA GTGGCCGGGT
901 TCGGGGAGAA GATCCACCAC TTGGATGACA TGCTCAAGAG CCAGCAGCGG
951 AAAGTCCGGC AAATGATAGA GCAGCTCCAG AATCAAAAG CTGTGATCCA
1001 GTCAAAGGAC GCCACCATCC AGGAGCTCAA GGAGAAAATC GCCTATCTGG
1051 AGGCAGAGAA TTTAGAGATG CATGACCGGA TGGAAACACCT GATAGAAAAA
1101 CAAATCAGTC ATGGCAACTT CAGCACCAGG GCCCGGGCCA AGACAGAGAA
1151 CCCGGGCAGT ATTAGGATAT CCAAGCCGCC TAGCCCGAAG CCCATGCCTG
1201 TCATCCGAGT GGTGGAAACC TGAGCTGCCT GGAGATGGTT GCTGCCATTG
1251 CTGCTGCCTC TGCCCTCGGAG AAGCCCACTG CCCCTGTGG CTGTAAACAC
1301 TGCCCTTTGAC TTCTGACTG TCCCCTGGCT GCACCCAGGA CTTGGGGCTC
1351 CTGTGTCTCA CCATTCCCAA GCCCCTGGCC ACTCTAAGCT GGGCAGACGG
1401 AGCAGGAGCA CCTATTCAAG GCACTGCAGC CCTTTGGAAG ACATTGTCTT
1451 GCAAGCAGGA GCCAGGGCAA TATCTATATT CCTACAGTGA CTATTTTCT
1501 CTGTAGAGAG CCTCCCTTCT GTTGTAGACT GGACTCTGGC TGCGCCATAA
1551 GCCAGGCCCT CATCAGATTG GGAGAGGTGA CAAGATTGTC CTCAGCCCTA
1601 AAAGCTGGAG ACACAGATGT CCAGAGTGAT TGGAGAAATG CCTGGGGGAA
1651 TGAAGTTCTT TCCACAAACA CAGCTCAGTT CTTAGCAACA AACTGTTTGT
1701 TTTTCTACTT GCTCCATCTG CAGCCTACGC TGCCCTGGCC TCCTGCAGAC
1751 AGATAGTGGG GTTACCTGGC AAGGCCTGGT GAGAGCCAGT GAACCTAAGC
1801 TTTGACTGGG TGGCCTTGTC TTTCTGGGGA GGAGGGAAATG TACATTACAG
1851 GAGTAGCCTT TTGCGGAAAA ATTCTTAGG GCTACAGACA GTCATGTGTG
1901 ACTTCTCTCT GCTGTGAAAA CTCCCAGAGT CTCTTTAGGG ATTTTCCCTA
1951 AGGTGTACCA CCAGGCACAC CTCAGTCTTC TTGACCCAGA GCCTGAAAC
2001 TGTTTTCACT GGGTTCCACC AGTCCCAGCA AAATCCTCTT TGTATTTATT
2051 TTGCTAAGTT ATTGGTGGTT TTGCTTACAT CTCATGATTG ATATAATACC
2101 AAAGTTCTAT AGCCTTCTCT TGCAGTATTG GGATTGTGTT GAAACCGGGA
2151 AAAGTTCTCC CATTAGGCTT GTTAATGTCA GAGTGACACT ATTATGAATC
2201 TTTCTCTCCC TTTCTCTCTG CTGTTCTTTC TCTCTTCTTC CTTCAAACCT
2251 GCTCTGCAGC TAAGGAAGGT GAGTCTACTT TCCCTGAGGC TTTGGGGTCA
2301 GAGTATATGT TGTGTGAGA AAGAGGGCAA TCAGGACTCT TCTGGGACCC
2351 AGATGAGTTC TTCCTAGACC CTTCTGAACC CCTTGCTCCA TAATTGGTCT
2401 TTTATCTCTG CTCTGAATGA CCCTGCAGGT CATCATGGTT TTCTTTTTTT
2451 ATTGTTTTTT TTTTTTCTG AGACAGAGTC TCACTCTGTC ACCCAGGCTG
2501 GAGTGCAGTG GCGGATCTC AGCTCACTGC AACCTCTGCC TCCCGGATTG
2551 AAGCGATTCT TCTGCCTCAG CCTCCCGAGT AGCTGGGACT ACAGGTGTGC
```



```

2601 CACCACGCCT GGCTGATTTT TGTATTTTGA GTAGAGATGG GGTTCACCA
2651 TACTGGCTAG GCTGGTCTCG AATTCCTGAC CTCAGGTGAT CCACCCACCT
2701 CGGCTTCCCA AAGTGCTAGG ATTATAGGCT TGAGCTACTG TGCCCGGCC
2751 ATGGTGTTTT TCTTTAGGGC TCTTCCTACA GCCTTGAGAA GTAGATAGGC
2801 ATCAGAGTAT GGTACTATAG GAATCAGAAA AATTCAAAAC AAATGTGGAT
2851 TAAGTGTTTA GGCTCTATGT GGCTCACGCA GCCAGAATCC TTAAGTCTGT
2901 GTGTTTCTGT GTCTCAAGAC TGGGCTCACA TTCTGGCTTT GTCCATAACA
2951 ATGCTCTGGG ATTTCAAGGA GTTCCCTCAT TTGTAAAATG AGGGGGTCAG
3001 AGCAGGTGAT ATCCATGTTT CTTCCTTTC TGATATGTT GTCTGTGGCA
3051 TATTCTTTGT ATGGCGAATT TAATAAATTA TATTAATGTG TCTAAAAAAA
3101 AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

98200312:

Tuftelin--aspects of protein and gene structure

97228909:

Timing of the expression of enamel gene products during mouse tooth development.

91340750:

Sequencing of bovine enamelin ("tuftelin") a novel acidic enamel protein.

Peptide information for frame 3

ORF from 51 bp to 1220 bp; peptide length: 390
 Category: strong similarity to known protein

```

1  MNGTRNWCTL VDVHPEDQAA GSVDIRLRTL QGELTGDELE HIAQKAGRKT
51 YAMVSSHAG HSLASELVES HDGHEEIKV YLKGSGDKM IHEKNINOLK
101 SEVQYIPEAR NCLQKLREDI SSKLDRNLGD SLHRQEIQVV LEKPNFGFSQS
151 PTALYSPPPE VDTICINEDVE SLRKTVDLL AKLQEAQRQH QSDCVAFEVT
201 LSRVQREAEQ SNVALQREED RVEQKEAEVG ELQRRLLGME TEHQALLAKV
251 REGEVALEEL RSNNADCOAE REKAATLEKE VAGLREKIH LDDMLKSQQR
301 KVRQMIEQLQ NSKAVIQSKD ATIQELKEKI AYLEAENLEM HDRMEHLIEK
351 QISHGNFSTQ ARAKTENPGS IRISKPPSPK PMPVIRVVET

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19g22, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_19g22, frame 3

Report for DKFZphut1_19g22.3

```

[LENGTH]      390
[MW]           44264.09
[pI]           5.68
[HOMOL]        TREMBL:AF047704_1 product: "tuftelin"; Mus musculus tuftelin mRNA, complete
cds. 0.0
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
2e-11
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 2e-11
[FUNCAT]       1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1643] 7e-11
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 1e-08
[FUNCAT]       03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]       03.13 meiosis [S. cerevisiae, YNL250w] 7e-08

```


[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 7e-08
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YKR095w] 1e-07
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 1e-05
 [FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YNL243w]
 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 -
 myosin-1 isoform] 4e-04
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 4e-04
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YMR294w] 7e-04
 [EC] 3.6.1.32 Myosin ATPase 8e-09
 [PIRKW] blocked amino end 1e-07
 [PIRKW] nucleus 1e-06
 [PIRKW] citrulline 1e-07
 [PIRKW] tandem repeat 8e-09
 [PIRKW] heterodimer 3e-06
 [PIRKW] DNA repair 2e-06
 [PIRKW] heart 8e-09
 [PIRKW] endocytosis 3e-07
 [PIRKW] transmembrane protein 4e-10
 [PIRKW] zinc finger 3e-07
 [PIRKW] metal binding 3e-07
 [PIRKW] muscle contraction 8e-09
 [PIRKW] acetylated amino end 1e-06
 [PIRKW] actin binding 8e-09
 [PIRKW] microtubule binding 1e-06
 [PIRKW] cell division control 1e-06
 [PIRKW] ATP 8e-09
 [PIRKW] chromosomal protein 3e-06
 [PIRKW] thick filament 8e-09
 [PIRKW] phosphoprotein 1e-145
 [PIRKW] skeletal muscle 8e-09
 [PIRKW] calcium binding 1e-07
 [PIRKW] meiosis 2e-06
 [PIRKW] alternative splicing 7e-08
 [PIRKW] DNA condensation 3e-06
 [PIRKW] coiled coil 4e-10
 [PIRKW] P-loop 8e-09
 [PIRKW] heptad repeat 1e-07
 [PIRKW] methylated amino acid 8e-09
 [PIRKW] immunoglobulin receptor 2e-06
 [PIRKW] peripheral membrane protein 3e-07
 [PIRKW] cardiac muscle 8e-09
 [PIRKW] hydrolase 8e-09
 [PIRKW] muscle 7e-08
 [PIRKW] EF hand 1e-07
 [PIRKW] cytoskeleton 7e-08
 [PIRKW] hair 1e-07
 [PIRKW] smooth muscle 7e-08
 [PIRKW] calmodulin binding 3e-07
 [SUPFAM] conserved hypothetical P115 protein 2e-09
 [SUPFAM] myosin heavy chain 8e-09
 [SUPFAM] RAD50 protein 2e-06
 [SUPFAM] calmodulin repeat homology 1e-07
 [SUPFAM] myosin motor domain homology 8e-09
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-06
 [SUPFAM] tropomyosin 7e-08
 [SUPFAM] protein-tyrosine kinase ret 3e-07
 [SUPFAM] plectin 1e-06
 [SUPFAM] trichohyalin 1e-07
 [SUPFAM] pleckstrin repeat homology 2e-06
 [SUPFAM] ribosomal protein S10 homology 1e-06
 [SUPFAM] protein kinase homology 3e-07
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
 [SUPFAM] giantin 4e-06
 [SUPFAM] kinesin-related protein KLP4 1e-06
 [SUPFAM] kinesin motor domain homology 1e-06
 [SUPFAM] human early endosome antigen 1 3e-07
 [SUPFAM] M5 protein 2e-06
 [PROSITE] MYRISTYL 1
 [PROSITE] AMIDATION 1
 [PROSITE] CK2_PHOSPHO_SITE 6


```

[PROSITE]      PKC_PHOSPHO_SITE      4
[PROSITE]      ASN_GLYCOSYLATION     2
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY         4.62 %
[KW]           COILED_COIL           35.13 %

```

```

SEQ      MNGTRNWCTLVDPEDQAAGSV DILRLTLQELTGDELEHIAQKAGRKTYAMVSSHSA
SEG      .....
PRD      cccccceeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....

SEQ      HSLASELVESHGHEEIIKVYLKGRSGDKMIHEKNINQLKSEVQYIQEARNCLQKLREDI
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....

SEQ      SSKLDRNLGDSLHRQEIQVVLEKPNGFSQSPTALYSSPPEVDTTCINEDVESLRKTVQDLL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....CCCCCCCCCCCCCCCC

SEQ      AKLQEAQRQHQSDCAVEFTLSRYQREAEQSNVALQREEDRVEQKEAEVGELQRRLLGME
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      TEHQALLAKVREGEVALEELRSNNADCAEREKAATLEKEVAGLREKIHHLDDMLKSQQR
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      KVRQMIEQLQNSKAVIQSKDATIQELKEKIAYLEAENLEMHDRMEHLIEKQISHGNFSTQ
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      ARAKTENPGSIRISKPPSPKPMFVIRVET
SEG      .....XXXXXXXXXXXXXXXXXXXXX...
PRD      hhccccccccceeeccccccccceeeccc
COILS    .....

```

Prosites for DKFZphut1_19g22.3

PS00001	2->6	ASN_GLYCOSYLATION	PDOC00001
PS00001	356->360	ASN_GLYCOSYLATION	PDOC00001
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	171->174	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00008	355->361	MYRISTYL	PDOC00008
PS00009	46->50	AMIDATION	PDOC00009

(No Pfam data available for DKFZphut1_19g22.3)

DKFZphut1_19h17

group: intracellular transport and trafficking

DKFZphut1_19h17 encodes a novel 879 amino acid protein, with similarity to *N.crassa* osbp oxysterol-binding protein.

The novel protein contains a oxysterol-binding protein family signature. Mammalian oxysterol-binding protein (OSBP) is a protein binds a variety of oxysterols (oxygenated derivatives of cholesterol). OSBP seems to play a complex role in the regulation of sterol metabolism. OSBP is a cytosolic/Golgi receptor for oxysterols such as 25-hydroxycholesterol, and thus a potential target of sphingomyelin turnover and cholesterol mobilization at the plasma membrane and/or Golgi apparatus. Therefore, the new protein seems to be involved in oxysterol metabolism.

The new protein can find application in modulating the response of cells to oxysterols. The protein can be used as marker for the golgi system. The Protein might be used to direct drugs to the golgi system in response to oxidative stress.

strong similarity to *C.elegans* ZK1086.1 and oxysterol-binding proteins

complete cDNA, complete cds, few EST hits
similarity to proteins involved in steroid biosynthesis

Sequenced by AGOWA

Locus: unknown

Insert length: 3828 bp

Poly A stretch at pos. 3811, polyadenylation signal at pos. 3784

```

1  GCGGCGCGCG CCGGCGCGCG CCGGAGCACCG AGCTCGCGCG ACGGTAGGAG
51  AAGCCCCCGA GCGCCCCACAG CATGAAGGAG GAGGCCTTCC TCCGCGCGCG
101 CTCTCTCCCTG TGTCACCTTT CCTCCACCCC TCAGAAAGTC GACCCCCGGA
151 AGCTCACCCG GAACCTTGCTC CTCAGCGGAG ACAATGAGCT CTACCCACTC
201 AGCCCCAGGA AGGACATGGA GCCCAACGGC CCGTCGCTGC CCAGGGATGA
251 AGGCCCCCGG ACCCAAGCT CTGCCACGAA GGTGCCACCG GCAGAGTACA
301 GGCTGTGCAA CGGGTCAGAC AAGGAATGTG TGTCCCCCAC CGCCAGGGTC
351 ACCAAGAAGG AGACTCTCAA GCGCAGAAAG GAGAACTACC GGCAGAGAGAA
401 GAAGCGCGCC ACACGGCAGC TGCTCAGCGC TCTGACAGAC CCCAGCGTGG
451 TCATCATGGC TGACAGCCTG AAGATCCGCG GCACCCTGAA GAGCTGGACC
501 AAGCTGTGGT GCGTGCTGAA GCCGGGGGTG CTGCTCATCT ACAAGACGCC
551 CAAGGTGGGG CAGTGGGTGG GCACGGTGCT GCTGCACTGC TGCGAGCTCA
601 TCGAGCGGCC CTCCAAGAAG GACGGCTTCT GCTTCAAGCT CTTCCACCCG
651 CTGGATCAGT CCGTCTGGGC CGTGAAGGGC CCCAAAGGTG AGAGCGTGGG
701 CTCATCACA CAGCCCCTGC CCAGCAGCTA CCTGATCTTC AGGGCCGCTC
751 CCGAGTCAGA TGGTCGCTGC TGGCTGGACG CCCTGGAGCT GGCCCTGCGC
801 TGCTCTAGCC TACTGAGACT GGGCACCTGC AAGCCGGGCG GAGACGGGGA
851 GCCAGGGACC TCGCCAGACG CATCACCTTC ATCGCTCTGT GGGCTGCCAG
901 CCTCAGCCAC TGTCCACCCA GACCAAGACC TGTTCCTACT GAACGGGTCT
951 TCCCTGGAGA ACGATGCATT CTCAGACAAG TCGGAGAGAG AGAACCTTGA
1001 GGAGTCAGAT ACCGAGACCC AGGACCATAG CCGGAAGACG GAGAGTGGCA
1051 GCGACCACTG AGAGACCCCT GGGGCCCCGG TCGGAGAGAG GACCACCTAT
1101 GTGGAGCAGG TCCAGGAGGA GCTGGGGGAG CTGGGCGAGG CGTCCCAGGT
1151 GGAGACAGTG TCAGAGGAGA ACAAGAGTCT GATGTGGACC CTGCTGAAGC
1201 AGCTACGGCC AGGCATGGAC CTGTCCCGCG TGGTGCTACC CACGTTCTGA
1251 CTGGAGCCGC GCTCCTTCTT GAACAAGCTC TCCGACTACT ACTACCACGC
1301 AGACCTGTCT TCCAGGGGCTG CGGTGGAGGA GGATGCCTAC AGCCCGATGA
1351 AGCTGGTGCT GCGGTGGTAC CTGTCTGGCT TCTACAAGAA GCCCAAGGGA
1401 ATCAAGAAGC CGTACAACCC CATCTGGGG GAGACCTTCC GCTGCTGCTG
1451 GTTCCACCCG CAGACTGACA GCCGCACATT CTACATAGCA GAGCAGGTGT
1501 CCCACCACCC GCCCGTGTCT GCCTTCCACG TCAGCAACCG GAAGGACGGC
1551 TTCTGCATCA GTGGCAGCAT CACAGCCAAG TCCAGGTTTT ATGGGAACCTC
1601 GCTGTCCGCG CTGCTGGACG GCAAAGCCAC GCTCACCTTC CTGAACCGAG
1651 CCGAGGATTA CACCCTTACC ATGCCCTACG CCCACTGCAA AGGAATCCTG
1701 TATGGCAGCA TGACCCTGGA GCTGGGTGGG AAGGTCACCA TCGAGTGTGC
1751 GAAGAACAAC TTCCAGGCCC AGCTGGAATT CAACTCAAG CCCTTCTTCG
1801 GGGGTAGCAC CAGCATCAAC CAGATCTCGG GAAAGATCAC GTCGGGAGAG
1851 GAAGTCTCTG CGAGCCTCAG TGGCCACTGG GACAGGGACG TGTATTATCAA
1901 GGAGGAAGGG AGCGGAAGCA GTGCGCTTTT CTGGACCCCG AGCGGGGAGG
1951 TCCGAGAGCA GAGGCTGAGG CAGCACACGG TGCCGCTGGA GGAGCAGACG
2001 GAGCTGGAGT CCGAGAGGCT CTGGCAGCAC GTCACCAGGG CCATCAGCAA
2051 GGGCCAGCAG CACAGGGCCA CACAGGAGAA GTTTGCACTG GAGGAGGCAC
2101 AGCGGCAGCG GGCCCGTGAG CGGCAGGAGA GCCTCATGCC CTGGAAGCCG
2151 CAGCTGTTCC ACCTGGACCC CATCACCCAG GAGTGGCACT ACCGATACGA
2201 GGACCACAGC CCCTGGGACC CCCTGAAGGA CATCGCCAGG TTTGAGCAAG
2251 ACGGGATCCT GCGGACCTTG CAGCAGGAGG CCGTGGCCCG CCAGACCACC

```



```

2301 TTCCTGGGCA GCCCAGGGCC CAGGCACGAG AGGTCTGGCC CAGACCAGCG
2351 GCTTCGCAAG GCCAGCGACC AGCCCTCCGG CCACAGCCAG GCCACGGAGA
2401 GCAGCGGATC CACGCCTGAG TCCTTGCCAG AGCTCTCAGA CGAGGAGCAG
2451 GATGGTGACT TTGTCCCTGG CGGTGAGAGC CCATGCCCTC GGTGCAGGAA
2501 GGAGGCGCGG CGGCTGCAGG CCCTGCACGA GGCATCCTC TCCATCCGAG
2551 AGGCCCAGCA GGAGCTGCAC AGGCACCTCT CGGCCATGCT GAGCTCCACG
2601 GCACGGGAG CACAGGCACC GACCCAGGCG CTCTGCAGA GCGCCCGATC
2651 CTGGTTCCCTG CTCTGCGTGT TCCTGGCGTG TCAGCTGTTT ATTAACCACA
2701 TCCTCAAATA GGAGCCCTGG GGGCAGAGCT CCTGGCCAGT CCCGAGCCCT
2751 CCCTCCCAGG CACCCAGCAC TTAAAGCCTG CTCCATGGAG GCAGAGAGGC
2801 CCGGCAAGCA CAGCCACTGT GACGGGGAGT CCAGGCGCAG GAGGGACCCG
2851 GGGCCACAAG GCGCTGCGGG CCCAGGTGTG CTGGGCCCTT CTCAGGGGCA
2901 CTGGCCTCTC TGCAGGGCCT TCCGCCCAGC GCTGGCCCTT ATGCTAAAGC
2951 CAAATGCAGC TTCTGCTGTG CGACGCACTC CTGGCCATCT TGCCGTGTCA
3001 CCCCCTGTCC GGCCTCCACT TGCCATGGGG GATGGATGGA TTTAGGTTGG
3051 GAGGGCCTGT GGGGGCCCTG GACAGTCACA CCCAGCAGC AGTGAGTGGG
3101 CAGGTTTGA GAGCAGCCA GGGAGCCCCG AGTGGCCAG GAGTCCCCC
3151 ACACACAGAT GCATAGGCCT GCCTTCCGGA GACCCTGTCC ACATTGCCGG
3201 GACCAACCTG GTGGGGCCAC TGGTGGGTGC CAGGGACAGG TTAGGGCCAC
3251 TCTGGGAAG GCATTTTGGT TTTTATTC ACCTCTGCT GTTGGATGG
3301 GAGCCCCACA GAGGCAGTCT CTGGAACCA CCCACCCCA CACCTGGACG
3351 CTCGCTCTGG TGGGGGCACA CGCAGGTGGA GGTGGTTGT GGTGCAGGTG
3401 TGTGCAGGGG TGTGGGGGGC GCAGGGGTGT GGCTTAGCTG GCGCCGACCC
3451 CAGGCCGGGG AGGCTCAAGT TCGCCACTTT ACTCAGACCG ATGCACAGTC
3501 TTCCCATTTT ACACCTTTTT AATAAACATA ATTGCAATAT TTTAGGTGGG
3551 CTGCGAGCTG CAGTCAGCCT TCACGTCTGG CCTCAGTCCC CGTGTCACTG
3601 CCGCTCTGCG TGTGCGTGTG CGCGTGTGT AGCCTCTACA CATATATATA
3651 TGTACAGAGC CTTAAACCAC ATCGTGGCGG TGCCGTCTGA GCTGTAGCGG
3701 GTGGCTTTGT TTCCAGTTTT TGTACCCGTG TCCTTGTCTC CCCTCCTCCC
3751 CCATCTGGGG ATGTGTCTGT GTTCCACACC TTGAAATAAA CAGACACATA
3801 CGTGTCTCT TAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

98315477:

The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes.

98146266:

A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

98146266:

A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

Peptide information for frame 3

ORF from 72 bp to 2708 bp; peptide length: 879
Category: strong similarity to known protein

```

1 MKEEAFRRR FSLCPPSSTP QKVDPRKLTR NLLSGDNEL YPLSPGKDME
51 PNGPSLPRDE GPPTPSSATK VPPAEYRLCN GSDKECVSPT ARVTKKETLK
101 AQKENYRQEK KRATRQLLSA LTDPSVVIMA DSLKIRGTLK SWTKLWCVLK
151 PGVLLIYKTP KVGQWVGTVL LHCCELIERP SKKDGFCFKL FHPLDQSVWA
201 VKGPKGESVG SITQPLPSY LIFRAASESD GRCWLDAL ELRCSSLLRL
251 GTCKPGRDGE PGTSPDASPS SLCGLPASAT VHPDQDLFPL NGSSLEND AF
301 SDKSERENPE ESDTETQDHS RKTESGSDQS ETPGAPVRRG TTYVEQVQEE
351 LGELGEASQV ETVSEENKSL MWTLKQLRP GMDLSRVVLP TFLVLEPRSF
401 NKLSDDYYHA DLLSRAAVEE DAYSRMKLV L RWYLSGFYKK PKGIKKPYNP
451 ILGETFRCCW FHPQTDSTF YIAEQVSHHP PVSAFHVSNR KDGFCISGSI
501 TAKSRFYGNS LSALLDGKAT LTFLNRAEDY TLTMPYAHCK GILYGMTLE
551 LGGKVTIECA KNNFQAQLEF KLKPFEGGST SINQISGKIT SGEEVLASLS
601 GHWRDVFVK EEGSGSSALF WTPSGEVRRQ RLRQHTVPLE EQTELESERL

```


651 WQHVTRAISK GDQHRATQEK FALEEAQRQR ARERQESLMP WKPQLFHLDP
 701 ITQEWHYRYE DHSFWDPLKD IAQFEQDGIL RTLQQEAVAR QTTLGSPGP
 751 RHERSGPDQR LRKASDQPSG HSQATESSGS TPESCPESLD EEQDGFVPG
 801 GESPCPRCRK EARRLQALHE AILSIREAQ ELHRHLSAML SSTARAAQAP
 851 TPGLLQSPRS WFLLCVFLAC QLFINHLK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19h17, frame 3

TREMBL:CEZK1086_2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid
 ZK1086, N = 1, Score = 1495, P = 2.7e-153

PIR:S25324 hypothetical protein YKR003w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 574, P = 8.5e-57

TREMBL:CEAF195_7 gene: "C32F10.1"; *Caenorhabditis elegans* cosmid
 C32F10., N = 1, Score = 588, P = 8.6e-57

PIR:S46796 hypothetical protein YKR003w homolog YHR001w - yeast
 (*Saccharomyces cerevisiae*), N = 1, Score = 585, P = 1.9e-56

TREMBL:NCOSBP_1 gene: "osbP"; product: "oxysterol-binding protein";
N.crassa mRNA for putative oxysterol-binding protein, N = 1, Score =
 571, P = 7e-55

TREMBL:AB017026_1 product: "oxysterol-binding protein"; *Mus musculus*
 mRNA for oxysterol-binding protein, complete cds., N = 2, Score = 328,
 P = 3e-35

>TREMBL:CEZK1086_2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid ZK1086
 Length = 751

HSPs:

Score = 1495 (224.3 bits), Expect = 2.7e-153, P = 2.7e-153
 Identities = 327/663 (49%), Positives = 430/663 (64%)

Query: 129 MADSLKIRGTLKSWTKLWCVLKPGLVLLIYKTPKV--GQWVGTVLLHCCELIERPSKKDGF 186
 MAD+LKIRG LK W + +CVLKPGL+L++YK K G WVGTVLL+ CELIERPSKKDGF
 Sbjct: 1 MADTLKIRGALKRWNRYCVLKPGLLLIYKHKKADRGDWVGTVLLNHCELIERSKKDGF 60

Query: 187 CFKLFHPLDQSVWAVKGPKEGVSIGT-QPLPSSYLIFRAASESDGRCWLDALALRCS 245
 CFKLFHFP+D S+W +GP G+S GS T PL +S+LI RA S+ GRCW+DALEL+ +C+
 Sbjct: 61 CFKLFHFPMDMSIWGNRGLQSGSFTLNPLNTSFLICRAPSDQAGRCWMDALELSFKCT 120

Query: 246 SLLRLGTCKPGRDGEPTGSPDASPSLLCGLPASATVHPDQDLFPLNGSSLENDAFSDK-S 304
 LL+ T D + G D+S + G + + D D G A S+ +
 Sbjct: 121 GLLKK-TMNE-LDDKNG---DSSMND--GQRDESRRMSRDS-----GDDTRELA VSETDA 168

Query: 305 ERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTT---YVEQVQEEGELGEASQVE 361
 E+ E D + +DH E G SET +R T ++ +E G G S E
 Sbjct: 169 EKHFEIDDVQEDH---EDGK-MSETSDT-IREAFTESAWIPSPKEVFGPDG--SLTE 220

Query: 362 TVSEENKSLMWTLLKQLRPGMDLSRVVLPFTVLEPRSFNLKLSDYHHADLLSRAAVEED 421
 V EENKSL+WTLLKQ+RPGMDLS+VVLPTF+LEPRSF KL+DYHHADL+S A E D
 Sbjct: 221 EVGEENKSLIWTLLKQIRPGMDLSKVVLPTFILEPRSFLEKLADYHHADLISEAVAEFD 280

Query: 422 AYSRMKLVLRWYLSGFYKKPKGIKKPYNPILGETFRCCWFHPQTDSTRTFYIAEQVSHHPP 481
 + R+ V +++LSGFYKKPKG+KKPYNPILGETFR C W HP S TFY+AEQVSHHPP
 Sbjct: 281 PFQIRVKVTKFFLSGFYKKPKGLKKPYNPILGETFRCKWEHPD-GSTTFYMAEQVSHHPP 339

Query: 482 VSAFHVSNRKDGFCISGSITAKSRFYGNLSALLDGKATLTLNRAEDYTLTMPYAHCKG 541
 VS+ ++NRK GF ISG+I AKS++YGNLSA+L GK LT LN E Y + +PYA+CKG
 Sbjct: 340 VSSLFITNRKAGFNISGTILAKSKYYGNLSAILAGKLRLTLNLGETYIVNLPYANCKG 399

Query: 542 ILYGTMTELEGGKVTIECAKNNFQAQLEFKLPFFGGSTSIHQISGKITSGEEVLASLSG 601
 I+ GTMT+ELGG+V IEC K ++ L+FKLP GG+ NQI G I G + LAS+ G
 Sbjct: 400 IMIGTMTMELGGEVNI ECEKTGYRTTLDKFKLPMGLGA--YNQIEGSIKYGSDRLASIEG 457

Query: 602 HWRDVFIFKEEGSGSSALFWTPSGEVRRQRLRQHTVPLEEQTELESERLWQHVTRAISKG 661
 WD + IK G W P+ EV + RL ++ + +EQ E ES +LW+HVT AIS
 Sbjct: 458 AWDGVIRIK--GPDGKKELWNPTPEVIKTRLPRIEINMDEQGEWESAKLWRHVTEAISNE 515

Query: 662 DQHRATQEKFALEEAQRQRARERQESLMPWKPQLFHLDPITQEWHYRYEDHSPWDPLKDI 721
 DQ++AT+EK ALE QR RA+ S +P + + F ++ Y + D+ PWD DI
 Sbjct: 516 DQYKATEEKTALENDQRARAK----SGIPHETKFFKKQH-GDDYVYIHADYRPWDNNNDI 570

Pedant information for DKFZphutel_19h17, frame 3

Report for DKFZphute1_19h17.3

```

SEQ      MKEEAFLLRRRFLCPPSSSTPQKVDPKILTRNLLLSGDNELYPLSPGKDMEPNGPSLPRDE
SEG
PRD      cccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
COILS
MEM      .....

SEQ      GPPTPSSATKVPPAEYRLCNGSDKECVSPTARVTKKETLKAQKENYRQEKKRATRQLLSA
SEG
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS
MEM      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      LTDPSPVIMADSLKIRGLKSWTKLWCVLKPGVLLIYKTPKVGQWVGTVLLHCCELIERP
SEG
PRD      hccccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS
MEM      CCC.....

SEQ      SKKDGFCFKLFHPLDQSVWAVGKPGESVGSITQPLPSSYLIFRAASESDGRCWLDALEL
SEG
PRD      cccccceeeccccccccccccccccccccccccccccccccccccceehhhhhhhhhhhhhhhhh
COILS
MEM      .....

SEQ      ALRCSSLLRLGTCKPGRDGEPTGSPDASPSSLCGLPASATVHPDQDLFLPLNGSSLENDAF
SEG
PRD      hhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
COILS
MEM      .....

SEQ      SDKSERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTTYEQVQEELGELGEASQV

```


Prosites for DKFZphute1_19h17.3

473

PS00005	301->304	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	455->458	PKC_PHOSPHO_SITE	PDOC00005
PS00005	488->491	PKC_PHOSPHO_SITE	PDOC00005
PS00005	501->504	PKC_PHOSPHO_SITE	PDOC00005
PS00005	586->589	PKC_PHOSPHO_SITE	PDOC00005
PS00005	647->650	PKC_PHOSPHO_SITE	PDOC00005
PS00005	824->827	PKC_PHOSPHO_SITE	PDOC00005
PS00005	843->846	PKC_PHOSPHO_SITE	PDOC00005
PS00005	857->860	PKC_PHOSPHO_SITE	PDOC00005
PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	362->366	CK2_PHOSPHO_SITE	PDOC00006
PS00006	590->594	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00006	659->663	CK2_PHOSPHO_SITE	PDOC00006
PS00006	713->717	CK2_PHOSPHO_SITE	PDOC00006
PS00006	755->759	CK2_PHOSPHO_SITE	PDOC00006
PS00006	780->784	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00006	789->793	CK2_PHOSPHO_SITE	PDOC00006
PS00006	824->828	CK2_PHOSPHO_SITE	PDOC00006
PS00007	402->409	TYR_PHOSPHO_SITE	PDOC00007
PS00007	415->424	TYR_PHOSPHO_SITE	PDOC00007
PS00008	137->143	MYRISTYL	PDOC00008
PS00008	163->169	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	326->332	MYRISTYL	PDOC00008
PS00008	381->387	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	541->547	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00008	728->734	MYRISTYL	PDOC00008
PS00013	860->871	PROKAR_LIPOPROTEIN	PDOC00013
PS01013	474->485	OSBP	PDOC00774

Pfam for DKFZphut1_19h17.3

HMM_NAME	PH (pleckstrin homology) domain		
HMM	*dvIREGWMyKWgswrkstgnWqrRWFvLrndpnzLiYYkddkdekPrYM		
Query	126	VVIMADSLKIRGTLKS---WTKLWCVLKP--GVLLIYKTP-KVGQWVG	167
HMM	lIdldcWrMidVEidWmmdndHCFiIWtrq.....		
Query	168	TVLLHCCELIERPSKGD---GFCFKLFHPLDQSVWAVKGPKEGKESVGSITO	214
HMMrtYYFQAeNeEEMmewMsaiRaiW*		
Query	215	PLPSSYLIFRAASESDGRCWLDALALR	243

DKFZphutel_19j11

group: uterus derived

DKFZphutel_19j11 encodes a novel 708 amino acid protein with C-terminal similarity to several known proteins, such as human KIAA0231 or murine ras binding protein Sur8.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

Strong similarity to KIAA0231, similarity to ras binding protein Sur8

EST AA854189 extends the sequence (294 Bp), with this sequence complete cDNA,

Sequenced by AGOWA

Locus: unknown

Insert length: 2343 bp

Poly A stretch at pos. 2323, polyadenylation signal at pos. 2295

```
1 GCTCCTGCTA ACCCCATCAC TGTGGAATG AAAGGCCTGA AGACAGATT
51 GGACCTTCAG CAGTACAGCT TTATAAATCA GATGTGTTAT GAGCGAGCCC
101 TCCACTGGTA TGCCAAGTAT TTCCTTACC TTGTCTCAT CCATACCCCTG
151 GTCTTTATGC TCTGCAGTAA CTTTGGTTC AAATCCCTG GTTCCAGCTC
201 CAAAATAGAA CATTTCATCT CCATTCTGGG GAAGTGTITT GACTCTCCTT
251 GGACCCACAG GGCTTTATCT GAAGTGTCTG GGGAGGACTC AGAAGAAAAG
301 GACAACAGGA AGAACACAT GAACAGGTCC AACACCATCC AATCTGGTCC
351 AGAAGGCAGC CTGGTCAACT CTCAGTCTTT AAAGTCCATT CCTGAGAAAT
401 TTGTAGTTGA TAAATCCACT GCAGGGGCTC TGGATAAAAA GGAAGTGCAG
451 CAGGCTAAGG CCTTATTTGA GAAGGTGAAG AAGTTCAGGC TGCATGTGGA
501 AGAAGGTGAT ATTCTATATG CCATGTATGT TCGCCAGACT GTACTTAAAG
551 TTATCAAATT CCTAATCATC ATTGCATATA ATAGTGCTCT GGTTCCTCAAG
601 GTCCAGTTTA CAGTGGACTG TAATGTGGAC ATTCAGGACA TGACTGGATA
651 TAAAAAATTT TCTTGCAATC ATACCATGGC ACACTTGTTT TCAAAATCTG
701 CCTTTTGCTA TCTGTGCTTT GTTAGTATCT ATGGATTGAC GTGCCTTTAT
751 ACCTTATACT GGCTGTTCTA CCGTTCCTTA CGGGAATATT CCTTTGAGTA
801 TGTCCGTCAG GAGACTGGAA TTGATGATAT TCCAGATGTG AAAAAAGACT
851 TTGCTTTTAT GCTTCATATG ATAGATCAGT ATGACCTCTC CTATTCCAAG
901 AGATTTCAGT TGTTCTGTG TGAAGTCAGT GAAAAACAAAT TAAAGCAGCT
951 GAACCTAAAT AACGAATGGA CTCTGATAAA ACTGAGGCAG AAGCTACAGA
1001 CAAATGCCCA TAATCGACTG GAATTGCCTC TTATCATGCT CTCTGGCCTT
1051 CCAGACACTG TTTTGAATAT CACAGAGTTG CAATCTCTAA AACTTGAAAT
1101 CATTAAAGAA GTAATGATAC CAGCCACCAT TGCACAGCTA GACAATCTTC
1151 AAGAGCTCTC TCTGCACCAG TGTTCTGTCA AAATCCACAG TCGGGCGCTC
1201 TCTTTCCTGA AGGAAAACCT CAAGGTCTTG AGCGTCAAGT TTGATGACAT
1251 GAGGGAACTC CCCCCCTGGA TGTATGGGCT CCGAAATCTG GAAGAGCTGT
1301 ACCTAGTTGG CTCTCTAAGT CATGATATTT CCAGAAATGT CACCCTTGAG
1351 TCTCTGCGGG ATCTCAAAAG CCTTAAATTT CTCTCTATCA AAAGCAACGT
1401 TTCCAAAATC CCTCAGGCAG TGGTTGATGT TTCCAGCCAT CTCCAGAAGA
1451 TGTGCATACA TAATGATGGC ACCAAGCTGG TGATGCTCAA CAACTTAAAG
1501 AAGATGACCA ATCTGACAGA GCTGGAGCTG GTCCACTGTG ACCTGGAGCG
1551 TATTCCTCAT GCTGTGTTCA GCCTACTCAG CCTCCAGGAA TTGGACCTGA
1601 AGGAAAACAA TCTGAAATCT ATAGAAGAAA TCGTTAGCTT TCAGCACTTA
1651 AGAAAGTTGA CAGTGCTAAA ACTGTGGCAT AACAGCATCA CCTACATCCC
1701 AGAGCATATA AAGAACTCA CCAGCCTGGA ACGCCTGTCC TTTAGTCACA
1751 ATAAAAATGA GGTGCTGCCT TCCCACCTCT TCCTATGCAA CAAGATCCGA
1801 TACTTGGACT TATCGTACAA TGACATTCTG TTTATCCCCC CTGAAATTGG
1851 AGTTCTACAA AGTTTACAGT ATTTTCCAT CACATGTAAC AAAGTGGAAA
1901 GCCTTCCAGA TGAACCTCTAC TTCTGCAAGA AACTTAAAC TCTGAAGATT
1951 GGAAAAAACA GCCTATCTGT ACTTTCACCG AAAATTGGAA ATTTGCTATT
2001 TCTTTCCTAC TTAGATGTAA AAGGTAATCA CTTTGAATC CTCCCTCCTG
2051 AACTGGGTGA CTGTGCGGCT CTGAAGCGAG CTGGTTTAGT TGTAGAAGAT
2101 GCTCTGTTTG AAACCTCTGCC TTCTGACGTC CGGGAGCAAA TGAACACAGA
2151 ATAACCTATT TTTCTGTTAA GTTTGACTGA AACACGCTTC TACCAATAC
2201 AGTATAAATA ATTAGGTAGT CTTAATGCCT TTCCTATTTT TTTTCCCTTT
2251 TCACACAAAA TGTACACAAA GATCGCGTAA GGAGTATGTA TTTTAAATAA
2301 AAATTTAATT GTATTTTTTC AATATTAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

96421675:
 Characterization of densin-180, a new brain-specific synaptic protein
 of the
 O-sialoglycoprotein family.

98337190:
 SUR-8, a conserved Ras-binding protein with leucine-rich
 repeats, positively regulates Ras-mediated signaling in *C.*
elegans.

Peptide information for frame 1

ORF from 28 bp to 2151 bp; peptide length: 708
 Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MKGLKTDLDL QQYSFINQMC YERALHWYAK YFPYLVLIHT LVFMLCSNFW
51 FKFFPGSSSKI EHFISILGKC FDSFWTTRAL SEVSGEDSEE KDNRRKNNMNR
101 SNTIQSGPEG SLVNSQSLKS IPEKFVVDKS TAGALDKKEG EQAKALFEKV
151 KKFRHLHVEEG DILYAMYVRQ TVLKVIKFLI IIAYNLSALVS KVQFTVDCNV
201 DIQDMTGYNK FSCNHTMAHL FSKLSFCYLC FVSIYGLTCL YTLVWLFYRS
251 LREYSFEYVR QETGIDDIPD VKNDFAFMLH MIDQYDPLYS KRFAVFLSEV
301 SENKQLQNLN NNEWTPDKLR QKLQTNNAHR LELPLIMLSG LPDTVFEITE
351 LQSLKLEIHK NVMIPTATIAQ LDNLQELSLH QCSVKIHSAA LSFLKENLKV
401 LSVKFDMDRE LPPWMYGLRN LEELYLVGSL SHDISRNVTL ESLRDLKSLK
451 ILSIKSNVSK IPQAVVDVSS HLQKMCIHND GTKLVMLNNL KKMTNLTELE
501 LVHCDLERIP HAVFSLSLQ ELDLKENNLK SIEEIVSFQH LRKLTVLKLV
551 HNSITYIPEH IKKLTSLERL SFSHNKIEVL PSHLFLCNKI RYLDLSYNDI
601 RFIPPEIGVL QSLQYFSITC NKVESLPDEL YFCKKLTCLK IGKNSLSVLS
651 PKIGNLLFLS YLDVKGNHFE ILPPELGDCR ALKRAGLVVE DALFETLPSD
701 VREQMKTE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19j11, frame 1

TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene,
 partial cds., N = 1, Score = 1408, P = 4.5e-144

TREMBL:AF054827_1 gene: "soc-2"; product: "leucine-rich repeat protein
 SOC-2"; *Caenorhabditis elegans* leucine-rich repeat protein SOC-2
 (soc-2) mRNA, complete cds., N = 1, Score = 304, P = 5.7e-24

TREMBL:RNU66707_1 product: "densin-180"; *Rattus norvegicus* densin-180
 mRNA, complete cds., N = 1, Score = 311, P = 7.4e-24

TREMBL:AF068921_1 product: "Ras-binding protein SUR-8"; *Mus musculus*
 Ras-binding protein SUR-8 mRNA, complete cds., N = 1, Score = 302, P =
 1.1e-23

>TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial
 cds.

Length = 476

HSPs:

Score = 1408 (211.3 bits), Expect = 4.5e-144, P = 4.5e-144
 Identities = 265/471 (56%), Positives = 361/471 (76%)

```

Query: 237 LTCLYTLVWLFYRSRLREYSFEYVRQETGIDDIPDVKNDFAFMLHMIDQYDPLYSKRFAVF 296
      LT Y+L+W+ SL++YSFE +R+++ DIPDVKNDFAF+LH+ DQYDPLYSKR++F
Sbjct: 1 LTSSYSLWMLRSSLKQYSFEALREKSNYSIDIPDVKNDFAFILHLADQYDPLYSKRFSIF 60

Query: 297 LSEVSENKQLQNLNNEWTPDKLRQKLQTNNAHRLELPLIMLSGLPDTVFEITELQSLKL 356
      LSEVSENKQLQ+NLNNEWTPDKLRQKLQTNNAHRLELPLIMLSGLPDTVFEITELQSLKL
Sbjct: 61 LSEVSENKQLQINLNNEWTPDKLRQKLQTNNAHRLELPLIMLSGLPDTVFEITELQSLKL 120

```


Query: 357 EIIKNVMIPATIAQLDNLQELSLHQCQSVKIHSAALSFLKENLKVLSVKFDDMRELPPWMY 416
 E+I V +P+ ++QL NL+EL ++ S+ + AL+FL+ENLK+L +KF +M ++P W++
 Sbjct: 121 ELIPEVKLPSAVSQLVNLKELRVYHSSLVVDHPALAFLEENLKILRLKFTMGKIPRWVF 180

Query: 417 GLRNLEELYLVGSLSHDISRNVLTLESRLDKSLKILSIKSNVSKIPQAVVDVSSHLOKMC 476
 L+NL+ELYL G + + + LE +DLK+L+ L +KS++S+IPQ V D+ LQK+
 Sbjct: 181 HLKNLKELYLSGCVLPEQLSTMQLLEGFQDLKLNRLTYLKSSLSRIPQVVTDLPLSLQKLS 240

Query: 477 IHNDGTKLVMLNNLKKMTNLTELELVHCDLERIPHAVFSLSLQELDLKENNLSIEEIV 536
 + N+G+KLV+LNNLKKM NL LEL+ CDLERIPH++FSL +L ELDL+ENNLK++EEI+
 Sbjct: 241 LONEGSKLVVLLNNLKKMVNLKSLELISCDLERIPHSTFSLNNLHEDLDRENNLKTVEEII 300

Query: 537 SFQHLRLKTLVLKLWHNSITYIPEHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLS 596
 SFQHL+ L+ LKLWHN+I YIP I L++LE+LS HN IE LP LFLC K+ YLDLS
 Sbjct: 301 SFQHLQLNSCLKLWHNNIAYIPAQIGALSNLQSLDHNNIENLPLQLFLCTKLHYLDLS 360

Query: 597 YNDIRFIPPEIGVLQSLQYFSITCNKVESLPDELYFCKKLTCLKIGKNSLSVLSPKIGNL 656
 YN + FIP EI L +LQYF++T N +E LPD L+ CKKL+ L +GKNSL LSP +G L
 Sbjct: 361 YNHLTFIPPEIQYLSNLQYFAVTNNNIEMLPDGLFQCKKLQCLLLGKNSLMNLSPHVGEL 420

Query: 657 LFLSYLDVKGNHFEILPPELGDCRALKRAGLVVEDALFETLPSDVREQMKT 707
 L++L++ GN+ E LPPEL C++LKR L+VE+ L TLP V E+++T
 Sbjct: 421 SNLTHLELIGNYLETLPELEGCSLKRNCILIVEENLLNTLPLPVTERLQT 471

Pedant information for DKFZphut1_19j11, frame 1

Report for DKFZphut1_19j11.1

[LENGTH] 708
 [MW] 81812.82
 [pI] 7.55
 [HOMOL] TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds.
 1e-149
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR353c] 3e-07
 [BLOCKS] BL00868F
 [BLOCKS] BL00985B Spermadhesins family proteins
 [EC] 3.4.17.3 Lysine carboxypeptidase 1e-08
 [EC] 4.6.1.1 Adenylate cyclase 3e-18
 [PIRKW] blocked amino end 1e-10
 [PIRKW] phosphotransferase 1e-09
 [PIRKW] nucleus 6e-08
 [PIRKW] duplication 3e-18
 [PIRKW] platelet 1e-10
 [PIRKW] tandem repeat 7e-16
 [PIRKW] keratan sulfate 7e-07
 [PIRKW] metallo-carboxypeptidase 1e-08
 [PIRKW] transmembrane protein 1e-10
 [PIRKW] serine/threonine-specific protein kinase 1e-09
 [PIRKW] autophosphorylation 1e-09
 [PIRKW] cartilage 7e-07
 [PIRKW] connective tissue 7e-07
 [PIRKW] magnesium 1e-09
 [PIRKW] cAMP biosynthesis 3e-18
 [PIRKW] ATP 1e-09
 [PIRKW] receptor 1e-09
 [PIRKW] leucine zipper 3e-13
 [PIRKW] glycoprotein 5e-12
 [PIRKW] extracellular matrix 7e-07
 [PIRKW] chondroitin sulfate proteoglycan 7e-07
 [PIRKW] cell adhesion 1e-08
 [PIRKW] hydrolase 1e-08
 [PIRKW] sulfoprotein 7e-07
 [PIRKW] membrane protein 1e-08
 [PIRKW] phosphorus-oxygen lyase 3e-18


```
SEQ      YLDVKGNHFEILPPELGDCRALKRAGLVVEDALFETLPSDVREQMKE
SEG      .....
PRD      hhhccccccccccccchhhhhhhhheeecccccccccccccccccc
MEM      .....
```

(No Pfam data available for DKFZphute1_19j11.1)

DKFZphutel_li2

group: transcription factor

DKFZphutel_li2 encodes a novel 594 amino acid protein similar to signal transducing proteins.

The protein contains 2 WD-40 repeats, which is typical for the beta-transducin subunit of G-proteins. In addition, the protein contains a C3HC4 zinc finger and a leucine zipper. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription.

The new protein can find application in modulating/blocking gene expression of genes controlled by this molecule.

similarity to Dictostelium myosin heavy chain kinase

complete cDNA, complete cds, EST hits
[PFAM] Zinc finger, C3HC4 type (RING finger)
[PFAM] WD domain, G-beta repeats
[SCOP] dltbgc_2.46.3.1.1 betal-subunit of the
signal-transducing G protei 3e-07

Sequenced by BMFZ

Locus: /map="16p13.3"

Insert length: 3584 bp

Poly A stretch at pos. 3555, polyadenylation signal at pos. 3537

```
1 GGGCGGGAGG TGCTTCCCAA GGACCGTAGA TGCCTCTCTA GAGCATGAGC
51 TCAGGGAAGA GTGCCCCGCTA CAACCGCTTC TCCGGGGGGC CCAGCAATCT
101 TCCCAACCCCA GACGTACCCA CAGGGACCAG AATGGAAACG ACCTTCGGAC
151 CCGCCTTTTC AGCCGTCACC ACCATCACA AAGCTGACGG GACCAGCACC
201 TACAAGCAGC ACTGCAGGAC AGCATGCCCC CCATCAGCAC TCCCCGCCGC
251 TCCGACTCCG CCATCTCTGT CCGCTCCCTG CACTCAGAGT CCAGCATGTC
301 TCTGCGCTCC ACATTCTCAC TGCCCGAGGA GGAGGAGGAG CCGGAGCCAC
351 TGGTGTTTGC GGAGCAGCCC TCGGTGAAGC TGTGCTGTCA GCTCTGCTGC
401 AGCGTCTTCA AAGACCCCGT GATCACCACG TGTGGGCACA CGTTCTGTAG
451 GAGATGCGCC TTGAAGTCAG AGAAGTGTC CGTGGAACAC GTCAAACCTGA
501 CCGTGGTGGT GAACAACATC GCGGTGGCCG AGCAGATCGG GGAGCTCTTC
551 ATCCACTGCG GGCACGGCTG CCGGGTAGCG GGCAGCGGGA AGCCCCCAT
601 CTTTGAGGTG GACCCCCGAG GGTGCCCTT CACCATCAAG CTCAGCGCCC
651 GGAAGGACCA CGAGGGCAGC TGTGACTACA GGCTGTGCG GTGTCCCAAC
701 AACCCAGCT GCCCCCGCT GCTCAGGATG AACCTGGAGG CCCACCTCAA
751 GGAGTGCAG CACATCAAAAT GCCCCCACTC CAAGTACGGG TGCACGTTCA
801 TCGGGAACCA GGACACTTAC GAGACCCACC TGGAGACTTG CCGCTTCGAG
851 GGCCTGAAGG AGTTTCTGCA GCAGACGGAT GACCGCTTCC ACGAGATGCA
901 CGTGGCTCTG GCCCAGAAGG ACCAGGAGAT CGCCTTCCGT CGCTCCATGC
951 TGGGAAAGCT CTCGGAGAAG ATCGACCAGC TAGAGAAGAG CCTGGAGCTC
1001 AAGTTTGACG TCCTGGACGA AAACCAGAGC AAGCTCAGCG AGGACCTCAT
1051 GGAGTTCGGG CGGGACGCAT CCATGTTAAA TGACGAGCTG TCCACATCA
1101 ACGCGCGGCT GAACATGGGC ATCCTAGGCT CCTACGACCC TCAGCAGATC
1151 TTCAAGTGCA AAGGGACCTT TGTGGGCCAC CAGGCCCCTG TGTGGTGTCT
1201 CTGCGCTTAC TCCATGGGTG ACCTGCTCTT CAGTGGCTCC TCTGACAAGA
1251 CCATCAAGGT GTGGGACACA TGTACCACCT ACAAGTGTCA GAAGACACTG
1301 GAGGGCCATG ATGGCATCGT GCTGGCTCTC TGCAATCCAGG GGTGCAAACT
1351 CTACAGCGCG TCTGCAGACT GCACCATCAT TGTGTGGGAC ATCCAGAACC
1401 TGCAGAAGGT GAACACCATC CGGGCCCATG ACAACCCGGT GTGCAGCTG
1451 GTCTCCTCAC ACAACGTGCT CTTACGCGGC TCCCTGAAGG CCATCAAGGT
1501 CTGGGACATC GTGGGCACTG AGCTGAAAGT GAAGAAGGAG CTCACAGGCC
1551 TCAACCACTG GGTGCGGGCC CTGGTGGCTG CCCAGAGCTA CCTGTACAGC
1601 GGCTCCTACC AGACAATCAA GATCTGGGAC ATCCGAACCC TTGACTGCAT
1651 CCACGTCTCT CAGACGTCTG GTGGCAGCGT CTAATCCATT GCTGTGACAA
1701 ATCACCACAT TGTCTGTGGC ACCTACGAGA ACCTCATCCA CGTGTGGGAC
1751 ATTGAGTCCA AGGAGCAGGT GCGGACCCCT ACGGGCCACG TGGGCACCGT
1801 GTATGCCCTG GCGGTCATCT CGACGCCAGA CCAGACCAAA GTCTTCAGTG
1851 CATCTACGA CCGGTCCCTC AGGGTCTGGA GTATGGACAA CATGATCTGC
1901 ACGCAGACCC TGCTGCGTCA CCAGGGCAGT GTCACCGCGC TGGCTGTGTC
1951 CCGGGGCCGA CTCTTCTCAG GGGCTGTGGA TAGCACTGTG AAGGTTTGA
2001 CTTGTCTACA GGATCCAGGC CAGGCTGTGG TTCCCTCTGA ACCAGCCCTG
2051 GACCTTTCTG AGCCAGGCTG GCCACATGGG GTGGTCTCGG GGTCTTCTGC
2101 TGCCCCGTGG GCATAGGTGG ACAGGCTCTG GCAGCCGGGC AGTGCCCTCC
2151 CCGTCCCATG CTCGGCGAGC CTCCCTCTAC TCGGCACTGT CCTTGTGCTC
2201 CAGCCCCCTT CTGGGTGCCA GGTACGACGC TTGCCCCGGC CCACCTTCCA
2251 TCCCCACCTT CCATCCCCAC CCTAGATGGA GCGAGGCGCT TTTTACTCAC
2301 CTTTCTTACC GTTTTATGAC TGTATGTAGA TTTGGTTACC TCCTGGTTGA
```



```

2351 AATAAATGCT CCACAGACTG TGGCTGTGAG TGGGGACAGC TCCTCGGGAC
2401 AAGGGGGCTG TGTGTGGCCT TGAGGTGGT GTGCACAGGC ACTGGCTGCT
2451 GTGAGTGGGG GGGCATGGGG CAGTTTCCTT TGGTGGACCC CAGGACTTCG
2501 GCCCACTCCG GGGCCTCCCC TCCCTGCTAG GAGGCAACTC GTCACACCCA
2551 AGCTGTCTGC CTCCAGTCCC ATCTCCCCCA ACACATGTGC CCCCCAAAAG
2601 TGAGCCAGGC ACCTCTGTTT CCTGTGTTT ATTGACAGCC GACGGCAGCG
2651 CCTTGCCCGC ACCTCCCCTG CCCACCTGCT GGAGCCCAGC CTGTGCCGCC
2701 CTCTGAGGAG AGGCCTGGGG GGACAGCTGG GCACGTCCAC TCGCAGGGAA
2751 ACACGGGGTG AGACAGCAGG AAGGGGCCCT GCACGCCGGG ACGCCACCTC
2801 CGCCAGCCGC CTCCACCCGC CCCACACCAC AATCGCTGGT TTTCGGCATT
2851 TTTTAAATTT TTTTTTAAAG AAACGTCAAA GTTGTGCCCA AACTGTGGA
2901 TCAGCAAACA CGATAGAGGA GACCAGTCAG TACTTCTTGG AGGGGCGAGG
2951 CCATGAGAGG AAAAGGGAGG GCGAGAATGA CCACACAACA CAGCCTTGGA
3001 CCATGAGCAG AAGCGTCCGT GGGAACTCCA CTGGGGTGGA TGGGCTGCCT
3051 GCACAGCCCC TGGAGAGGGG GCCAGGCACA CCCTCAGAGG AGCTGCAAGC
3101 CCGTGGCCTG GCCTGCTACA TGCCCTGCTT CCACGTGGCT GCCACGCTGA
3151 CACACCCACA TTCACCAAAC CCACCCGCGC CCTGGGACGC AGCCACGCCA
3201 GGAGGAGGAC ACGGCCGCGC AGAGCAAGGC ACAACCTCGA GTTCTTGGGG
3251 CGCAGAGAAC TTAGGAGAGA AGCACGGAGG AGCCCCGGC AGAGCACCCG
3301 CCCCCGGGCC CCAGCCTTCC ACCTGTGCTA GCAGCCTGGG GCCTCCACTC
3351 TGGCCGGAGG AAGGACCGCA GGCAGACAGC CTGGGCCTCT AACAGCTTTT
3401 GTCCGGAGCT AGACTTCGTG TCCTTTCAGT TGGTAAATGG TTTTCTATAG
3451 AATCAATAAT ATTTCTTTCT TTAAATATAT ATTTGTTAAA GTTATACCTT
3501 TTTGTTTCTC TGGGGAAATC CGCCTCAGCT CATTCCCAAT AAATTAATAC
3551 TCTTGATAAA AAAAAAAAAA AGAAAAAAAA AAAA

```

BLAST Results

Entry HSBE from database EMBL:
Homo sapiens (clone exon trap d5) chromosome 16p13.3 gene, exon.
Score = 2375, P = 7.1e-101, identities = 475/475

Entry HSBD from database EMBL:
Homo sapiens (clone exon trap d32) chromosome 16p13.3 gene, exon.
Score = 876, P = 3.0e-31, identities = 176/177

Medline entries

95122486:
Structural analysis of myosin heavy chain kinase A from Dictyostelium. Evidence for a highly divergent protein kinase domain, an amino-terminal coiled-coil domain, and a domain homologous to the beta-subunit of heterotrimeric G proteins.

96149460:
Dictyostelium myosin heavy chain kinase A regulates myosin localization during growth and development.

97277316:
Identification of a protein kinase from Dictyostelium with homology to the novel catalytic domain of myosin heavy chain kinase A.

96009891:
A gene responsible for vegetative incompatibility in the fungus Podospora anserina encodes a protein with a GTP-binding motif and G beta homologous domain.

Peptide information for frame 2

ORF from 224 bp to 2005 bp; peptide length: 594
Category: similarity to known protein
Prosites motifs: ZINC_FINGER_C3HC4 (70-80)
LEUCINE_ZIPPER (436-458)
LEUCINE_ZIPPER (436-458)
G_BETA_REPEATS (335-355)
G_BETA_REPEATS (376-391)


```

1 MPPISTPRRS DSAISVRSLSH SESSMSLRST FSLPEEEEEEP EPLVFAEQPS
51 VKLCCQLCCS VFKDPVITTC GHTFCRRCAL KSEKCPVDNV KLTVVVNNIA
101 VAEQIGELFI HCRHGCRVAG SGKPPIFEVD PRGCPFTIKL SARKDHEGSC
151 DYRPVRCFNN PSCPPLLRMN LEAHLKECEH IKCPHSHKYGC TFIGNQDTYE
201 THLETCTREFG LKEFLQQTDD RFHEMHVALA QKDQEI AFLR SMLGKLSEKI
251 DQLEKSLELK FDLVDENQSK LSEDLMEFRR DASMLNDELS HINARLNMGI
301 LGSYDPQQIF KCKGTFVGHQ GPVWCLCVYS MGDLLFSGSS DKTIKVWDTG
351 TTYKCQKTLE GHGIVLALC IQGCKLYSGS ADCTIIVWDI QNLQKVNTIR
401 AHDNPVCTLV SSHNVLFSGS LKAIKVWDIV GTLKLKKEK TGLNHWVRAL
451 VAAQSYLYSG SYQTIKIWDI RTLDCHIHLQ TSGGSVYSIA VTNHHIVCGT
501 YENLIHVWDI ESKEQVRTLT GHVGTVYALA VISTPDQTKV FSASYDRSLR
551 VWSMDNMICT QTLRHQGSV TALAVSRGRL FSGAVDSTVK VWTG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphutel_li2, frame 2

SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B)., N = 1, Score = 419, P = 3.6e-37

SWISSPROT:HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1., N = 1, Score = 392, P = 3.1e-33

SWISSPROT:YDJ5_SCHPO HYPOTHETICAL 67.1 KD TRP-ASP REPEATS CONTAINING PROTEIN C57A10.05C IN CHROMOSOME I., N = 1, Score = 357, P = 4.1e-30

TREMBL:AF032878.1 gene: "slimb"; product: "slimb"; Drosophila melanogaster Slimb (slimb) mRNA, complete cds., N = 1, Score = 347, P = 1.7e-29

>SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B).
Length = 732

HSPs:

Score = 419 (62.9 bits), Expect = 3.6e-37, P = 3.6e-37
Identities = 96/268 (35%), Positives = 158/268 (58%)

```

Query:   325 CLCVYSMGDLLFSGSSDKTIKVWD-TCTTYKCQKTLEGHGIVLALCIQGCKLYSGSADC 383
          C+C   +LLF+G SD +I+V+D   +C +TL+GH+G V ++C   L+SGS+D
Sbjct:   467 CIC----DNLLFTGCSDNSIRVVDYKSNMECVQTLKGHEGPVESICYNDQYLFSGSSDH 522

Query:   384 TIIVWDIQNLQKVNTIRAHNDNPVCTLVSSHNVLFSGSL-KAIKVWDIVGTLEKLKKELTG 442
          +I VWD++ L+ + T+ HD PV T++ + LFSGS K IKVWD+   L+ K L
Sbjct:   523 SIKVWDLKKLRICFTLEGHDKPVHTVLLNDKYLFGSGSSDKTIKVWDL--KTLECKYTLES 580

Query:   443 LNHWVRALVAAQSYLYSGSY-QTIKIWDI RTLDCHIHLQ TSGGSVYSIAVTNHHIVCGTY 501
          V+ L + YL+SGS +TIK+WD++T C + L+   V +I +   ++ G+Y
Sbjct:   581 HARAVKTL CISGQYLFSGSNDKTIKVWDLKTFRCNYTLKGHTKWVTTCILGTNLNLSGSY 640

Query:   502 ENLIHVWDIESKEQVRTLTGHVGTVYALAVISTPDQTKVFSASYDRSLRVWSMDNMICTQ 561
          + I VW++S E TL GH V + + D+ +F+AS D +++W ++ + C
Sbjct:   641 DKTIRVWNLKSLECSATLRGHRWVEHVMVIC---DKL-LFTASDDNTIKIWDLETLCRNT 696

Query:   562 TLLRHQGSVTALAVSRGR--LFSGAVDSTVKVW 592
          TL H +V LAV + + S + D +++VW
Sbjct:   697 TLEGHNATVQCLAVWEDKKCVISCSHDQSIRVW 729

```

Score = 415 (62.3 bits), Expect = 1.2e-36, P = 1.2e-36
Identities = 113/303 (37%), Positives = 166/303 (54%)

```

Query:   255 KSLEL-KFDVLDENQSKLSEDLMEFRRDASMLNDEL-SHINARLNMGILGS-----YD 305
          KS++L K ++L N+ K S +L + ++ + SH+ N+ G YD
Sbjct:   427 KSIDLEKPEILINNKKESINLETIKIYHVTSHLCICDNLLFTGCSDNSIRVVD 486

Query:   306 -PQQIFKCKGTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVWD-TCTTYKCQKTLEGHG 364
          Q +C T GH+GPV +C Y+ LFSGSSD+IKVWD +C TLEGHG
Sbjct:   487 YKSNMECVQTLKGHEGPVESIC-YN-DQYLFSGSSDHSIKVWDL-KKLRICFTLEGHG 543

Query:   365 IVALCIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHNDNPVCTLVSSHNVLFSGSL-KA 423
          V + + L+SGS+D TI VWD++ L+ T+ +H V TL S LFSGS K
Sbjct:   544 PVHTVLLNDKYLFGSGSSDKTIKVWDLKTLECKYTLESHARAVKTL CISGQYLFSGSNDKT 603

Query:   424 IKVWDIVGTLEKLKKELTGLNHWVRALVAAQSYLYSGSY-QTIKIWDI RTLDCHIHLQTS 482
          IKVWD+ + L G WV + + LYSGSY +TI++W++++L+C L+
Sbjct:   604 IKVWDL--KTFRCNYTLKGHTKWVTTCILGTNLNLSGSYDKTIRVWNLKSLECSATLRGH 661

```


Query: 483 GGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVYALAVISTPDQTKVFS 542
 V + + + + + N I +WD+E+ TL GH TV LAV D+ V S
 Sbjct: 662 DRWVEHMMVICDKLLFTASDDNTIKIWDLETLCRNTTLEGHNATVQCLAVWE--DKKCVIS 719

Query: 543 ASYDRSLRVW 552
 S+D+S+RVW
 Sbjct: 720 CSHDQSIRVW 729

Score = 262 (39.3 bits), Expect = 3.2e-19, P = 3.2e-19
 Identities = 60/184 (32%), Positives = 109/184 (59%)

Query: 352 TYKQKRTLEGHGDIIVLALCIQCGKLYSGSADCTIIVWDI--QNLQKVNTIRAHNDNPVCTL 409
 T K +T++G+ + LCI L++G +D +I V+D QN++ V T++ H+ PV ++
 Sbjct: 450 TIKLIETIKGYH-VTSHLCICDNLFTGCSDNSIRVYDKSQNMECVQTLKGHEGPVESI 508

Query: 410 VSSHNVLFSGSLK-AIKVWDIVGTTELKKELTGLNHWVRALVAAQSYLYSGSY-QTIKI 467
 + LFSGS +IKVWD+ +L+ L G + V ++ YL+SGS +TIK+
 Sbjct: 509 CYNDQYLFSGSSDHSIKVWDL--KKLRCIFTLEGHDKPVHTVLLNDKYLFSGSSDKTIKV 566

Query: 468 WDRTLDLCIHVLQTSGGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVY 527
 WD++TL+C + L++ +V ++ ++ ++ G+ + I VWD+++ TL GH V
 Sbjct: 567 WDLKTECKYTLESHARAVKTLCSIGQYLFSGSNDKTIKVWDLKTRCNYTLKGHTKWT 626

Query: 528 ALAVIST 534
 + ++ T
 Sbjct: 627 TICILGT 633

Score = 173 (26.0 bits), Expect = 1.7e-09, P = 1.7e-09
 Identities = 43/118 (36%), Positives = 65/118 (55%)

Query: 310 FKCKGTFFVGHQGPVWCLCVYSMGDLFSGSSDKTIKVWDTCTTYKQKRTLEGHGDIIVLAL 369
 F+C T GH V +C+ +G L+SGS DKTIVW+ + +C TL GHD V +
 Sbjct: 612 FRCNYTLKGHTKWTTCICI--LGTNLYSGSYDKTIRVWNL-KSLECSATLRGHRWVEHM 668

Query: 370 CIQCGKLYSGSADCTIIVWDIQLQKVNTIRAHNDNPV-CTLVSSH--VLFSGSLKAIKV 426
 I L++ S D TI +WD++ L+ T+ H+ V C V V+ ++I+V
 Sbjct: 669 VICDKLLFTASDDNTIKIWDLETLCRNTTLEGHNATVQCLAVWEDKKCVISCSHDQSIRV 728

Query: 427 W 427
 W
 Sbjct: 729 W 729

Pedant information for DKFZphutel_1i2, frame 2

 Report for DKFZphutel_1i2.2

[LENGTH] 594
 [MW] 66541.94
 [pI] 6.64
 [HOMOL] SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B). 3e-37

[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YCR072c beta-transducin family] 2e-15
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 3e-11
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 8e-09
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-07
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 5e-07

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 3e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 2e-04
 [FUNCAT] 01.03.07 deoxyribonucleotide metabolism [S. cerevisiae, YOR269w] 2e-04
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR212w] 0.001
 [BLOCKS] BL00678
 [BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
 [SCOP] dltbgsd_2.46.3.1.1 beta1-subunit of the signal-transducing 3e-10
 [EC] 2.7.1.129 Myosin-heavy-chain kinase 3e-26
 [PIRKW] phosphotransferase 3e-26
 [PIRKW] nucleus 1e-06
 [PIRKW] plasma 9e-08
 [PIRKW] duplication 3e-25
 [PIRKW] hormone 9e-08
 [PIRKW] zinc 3e-09
 [PIRKW] cell cycle control 4e-13
 [PIRKW] transmembrane protein 3e-12
 [PIRKW] zinc finger 1e-08
 [PIRKW] stomach 9e-08
 [PIRKW] DNA binding 9e-06
 [PIRKW] autophosphorylation 3e-26
 [PIRKW] phosphoprotein 3e-26
 [PIRKW] signal transduction 5e-08
 [PIRKW] heterotrimer 5e-08
 [PIRKW] coiled coil 3e-26
 [PIRKW] multimer 3e-26
 [PIRKW] transcription regulation 4e-10
 [PIRKW] GTP binding 5e-08
 [SUPFAM] chromobox homology 9e-06
 [SUPFAM] RING finger homology 3e-09
 [SUPFAM] coatomer complex beta' chain 1e-07
 [SUPFAM] WD repeat homology 3e-26
 [SUPFAM] yeast coatomer complex alpha chain 3e-12
 [SUPFAM] GTP-binding regulatory protein beta chain 5e-08
 [SUPFAM] PRL1 protein 2e-09
 [PROSITE] WD_REPEATS 2
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 14
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] ZINC_FINGER_C3HC4 1
 [PROSITE] PKC_PHOSPHO_SITE 18
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 6.23 %
 [KW] COILED_COIL 6.73 %

SEQ MPPISTPRRSDSAISVRSLSHSESSMSLRSTFSLPEEEEEPEPLVFAEQPSVKLCCQLCCS
 SEGXXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXX.....
 COILS
 1gg2B
 SEQ VFKDPVITTCGHTFCRRALKSEKCPVDNVKLTVVVNNIAVAEQIGELFIHCRHGCRVAG
 SEG
 COILS
 1gg2B
 SEQ SGKPPIFEVDPRGCPFTIKLSARKDHEGSCDYRPRVPCPNPSCPLLRLMNEAHLKECEH
 SEG
 COILS
 1gg2B
 SEQ IKCPHISKYGCTFIGNQDTYETHLETCTRFEGLEKFLQQTDDRFHEMVALAQKDQEI AFLR
 SEG
 COILSCCCCCCCCCCCCCCCC
 1gg2B
 SEQ SMLGKLSKIDQLEKSLKFDVLDENQSKLSLMEFRDASMLNDELSHINARLNMGI
 SEG
 COILS CCCCCCCCCCCCCCCCCCCCCCCCCC.....
 1gg2B
 SEQ LGSYDPQQIFKCKGTFVGHQGPVWCLCVYSMDLLFSGSSDKTIKVWDTCTTYKQKLTLE
 SEG
 COILSEECCCCCEEEEEETTTTCEEEEEETTTTEEEEEEG-GGCEEEEEEE
 1gg2B


```

SEQ      GHDGIVLALCIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHDNFVCTLVSSHNVLFSGS
SEG      .....
COILS    .....
lgg2B    CCCCCEEEEETTCEEEEEETTTCEEEEEETTTTEEEEE-CTTTTCCCEEE.....

SEQ      LKAIKVWDIVGTELKCLKELTGLNHWVRALVAAQSYLYSGSYQTIKIWDIRTLDCIHVLQ
SEG      .....XXXXXXXXXXXXX.....
COILS    .....
lgg2B    .....

SEQ      TSGGSVYSIAVTNNHHIVCGTYENLIHVMDIESKEQVRTLTGHVGTVYALAVISTPDQTKV
SEG      .....
COILS    .....
lgg2B    .....

SEQ      FSASYDRSLRVWMDNMICTQTLLRHQGSVTALAVSRGRLFSGAVDSTVKVWTC
SEG      .....
COILS    .....
lgg2B    .....

```

Prosites for DKFZphute1_li2.2

PS00001	267->271	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	50->53	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	141->144	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	340->343	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	464->467	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	330->334	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	194->200	MYRISTYL	PDOC00008
PS00008	299->305	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	364->370	MYRISTYL	PDOC00008
PS00008	379->385	MYRISTYL	PDOC00008
PS00008	419->425	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	484->490	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	568->574	MYRISTYL	PDOC00008
PS00008	583->589	MYRISTYL	PDOC00008
PS00518	70->80	ZINC_FINGER_C3HC4	PDOC00449
PS00029	436->458	LEUCINE_ZIPPER	PDOC00029
PS00678	335->350	WD_REPEATS	PDOC00574
PS00678	376->391	WD_REPEATS	PDOC00574

Pfam for DKFZphute1_li2.2

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVWCVaFSPDGrWFiVSGSWDgTCRLWD*
 ++GH ++VWC+ + G + ++SGS D+T+++WD

Query 316 FVGHQGPVWCLCVYSMGDL-LFSGSSDKTIKVWD 348

22.93 519 553 1 34 dkfzphute1_li2.2 similarity to Dictostelium myosin heavy chain kinase

Alignment to HMM consensus:

Query *MrGHnnWVWCVaF..SPDGzWFIvSGSWDgTCRLWD*
 ++GH ++V+++A+ +PD ++S+S D+++R+W+
 dkfzphutel 519 LTGHVGTVYALAVISTPDQTK-VFSASYDRSLRVWS 553

HMM_NAME Zinc finger, C3HC4 type (RING finger)

HMM *CPICFcTFQldyPWPfdePmMlPCgHsFCypCIrrW..CPmC*
 C++C + F++P++++CGH+FC+ C +++ CP+
 Query 55 CQLC-----CSV---FKDPVITTCGHTFCRRCALKSEKCPVD 88

DKFZphutel_20b19

group: metabolism

DKFZphutel_20b19 encodes a novel 486 amino acid protein with similarity to bacterial sarcosine oxidases (EC 1.5.3.1.)

The novel protein seems to be a novel enzyme with sarcosine oxidase activity.

The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

similarity to sarcosine oxidases

membrane regions: 1

Summary DKFZphutel_20b19 encodes a novel 486 amino acid protein, with similarity to sarcosine oxidases.

similarity to sarcosine oxidases

complete cDNA?, complete cds potential start at Bp 48, EST hits,

Sequenced by AGOWA

Locus: unknown

Insert length: 1967 bp

Poly A stretch at pos. 1950, no polyadenylation signal found

```
1 AGCGAGGCGAG CAGTGCAGCT TTCAGAGGGT CCGGGCTCAG AGGGGTTATG
51 ATTCGGAGGG TTCTGCCGCA CGGCATGGGC CGGGGCTCT TGACCCGGAG
101 GCCAGGCACG CGCAGAGGAG GCTTTCTCT GGAAGGTTGT GGAAGGTTGT
151 CTGAGATTAA GAAGAAGATC AAGTCGATCC TGCCTGGAAG GTCCTGTGAT
201 CTAAGTCAAG ACACCAAGCA CCTGCCTCCC GAGCACTCGG ATGTGGTGAT
251 CGTGGGAGGT GGGGTGCTTG GCTTGTCTGT GGCCTATTGG CTGAAGAAGC
301 TGGAGAGCAG ACGAGGTGCT ATTCGAGTGC TAGTGGTGGG ACGGACCAC
351 ACGTATTAC AGGCCTCCAC TGGGCTCTCA GTAGGTGGGA TTTGTCAGCA
401 GTTCTCATTT CCTGAGAAC TCCAGCTCTC CCTCTTTTCA GCCAGCTTTC
451 TACGGGAACAT CAATGAGTAC CTGGCCGTAG TCGATGCTCC TCCCTGGGAC
501 CTCGGGTTC AACCCTCGGG CTACCTCTTG CTGGCTTCAG AAAAGGATGC
551 TGCAGCCATG GAGAGCAACG TGAAGTGCA GAGGCAGGAG GGAGCCAAAG
601 TTTCTCTGAT GTCTCTGAT CAGCTTCGGA ACAAGTTTCC CTGGATAAAG
651 ACAGAGGGAG TGGCTTTGGC GTCTTATGGG ATGGAGGACG AAGGTTGGTT
701 TGACCCCTGG TGTCTGCTCC AGGGGCTTCG GCGAAAGGTC CAGTCCTTGG
751 GAGTCTTTT CTGCCAGGGA GAGGTGACAC GTTTTGTCTC TTCATCTCAA
801 CGCATGTTGA CCACAGATGA CAAAGCGGTG GTCTTGAAAA GGATCCATGA
851 AGTCCATGTG AAGATGGACC GCAGCCTGGA GTACCAGCCT GTGGAATGCG
901 CCATTGTGAT CAACGCAGCC GGAGCCTGGT CTGCGCAAAT CGCAGCACTG
951 GCTGGTGTG GAGAGGGGCC GCCTGGCACC CTGCAGGGCA CCAAGCTACC
1001 TGTGGAGCCG AGGAAAAGGT ATGTGTATGT GTGGCACTGC CCCCAGGGAC
1051 CAGGCTTAGA GACTCCGCTT GTTGCAGACA CCAGTGGAGC CTATTTTCGC
1101 CGGGAAGGAT TAGGTAGCAA CTACCTAGGT GGTCTGAGCC CCACTGAGCA
1151 GGAAGAACC GACCCGCGCA ACCTGGAAGT GGACCATGAT TTCTTCCAGG
1201 ACAAGGTGTG GCCCCATTG GCCCTGAGGG TCCCAGCTTT TGAGACTCTG
1251 AAGGTTTACA GCGCCTGGGC CGGCTATTAC GACTACAACA CCTTTGACCA
1301 GAATGGCGTG GTGGGCCCC ACCCGCTAGT TGTCAACATG TACTTTGCTA
1351 CTGGCTTCAG TGGTACAGGG CTCCAGCAGG CCCCTGGCAT TGGGCGAGCT
1401 GTAGCAGAGA TGGTACTGAA GGGCAGGTT CAGACCATCG ACCTGAGCCC
1451 CTTCCTCTTT ACCCGCTTTT ACTTGGGAGA GAAGATCCAG GAGAACAACA
1501 TCATCTGAGC ATGTGTGCTC TGCACTGGCT CCACTGGCTT GCATCCTGGC
1551 TGTGTTTACA GCCTTGTGTT CTGCTTCCAT CTCCCCAGT ACTGTGCCAG
1601 GCCTTCTCCC CCTCCCCAGT GTCCTCTCCT CTCAGGCAGG CCATTGCACC
1651 CATATGGCTG GGCAGGCACA GGCAGTGAGG CCGAGGCCAA TAGCGAGTGA
1701 TGAGCGGGAT CCTAGGACTG ATCTGTAGCC CATGCTGATG TCACCCACCA
1751 GGGCAATCCA TCTGAGGCC TGAGCACCTT GGGCCAGGAC TGGCTTCATC
1801 CTGGCACTGA CCAGGAAAGA CTGCTCTGA CCCTCTTAGC AGACAGAGCC
1851 CAGGCATGGG AGCACTCTGG GGCAGCCTGG CTCAGGTTTA TTGATTTTCG
1901 TCTGTTTACC CTATCCATTA ATCAATACAT GTAATTAAC CTCTCCCTCC
1951 AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 48 bp to 1505 bp; peptide length: 486
 Category: similarity to known protein

```

1 MIRRVLPHGM GRGLLTRRPG TRRGGSFSLDW DGKVSEIKKK IKSILPGRSC
51 DLLQDTSHLP PEHSDVVIIVG GGVGLSVAY WLKKLESRRG AIRVLVVERD
101 HTYSQASTGL SVGGICQQFS LPENIQLSLF SASFLRNINE YLAVVDAPPL
151 DLRFNPSGYL LLASEKDAAA MESNVKVQRQ EGAKVSLMSP DQLRNKFPWI
201 NTEGVALASY GMEDEGWFDW WCLLQGLRRK VQSLGVLFCQ GEVTRFVSSS
251 QRMLTTDDKA VVLKRIHEVH VKMDRSLEYQ PVECAIVINA AGAWSAQIAA
301 LAGVGEGPPG TLQGTKLPEV PRKRYVYVWH CPQGPGLTLP LVADTSGAYF
351 RREGLGSNYL GGRSPTEQEE PDPANLEVDH DFFQDKVWPH LALRVPAFET
401 LKVSQAWAGY YDNTFDQNG VVGPHPLVWN MYFATGFSGH GLQQAPGIGR
451 AVAEMVLKGR FQTIDLSPLF FTRFYLGEKI QENNII

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_20b19, frame 3

TREMBL:CEM04B2_4 gene: "M04B2.4"; *Caenorhabditis elegans* cosmid M04B2,
 N = 1, Score = 801, P = 9.2e-80

PIR:B71184 probable sarcosine oxidase - *Pyrococcus horikoshii*, N = 2,
 Score = 194, P = 2e-26

PIR:B69284 sarcosine oxidase, subunit beta (soxB) homolog -
Archaeoglobus fulgidus, N = 3, Score = 189, P = 8.2e-22

TREMBL:AF042732_1 gene: "Bb"; product: "unknown protein"; *Anopheles gambiae* (Bb) gene, partial cds; and TU37B2 (TU37B2) and diphenol oxidase-A2 (Dox-A2) genes, complete cds., N = 1, Score = 386, P = 8.7e-36

PIR:F71008 probable sarcosine oxidase - *Pyrococcus horikoshii*, N = 2,
 Score = 200, P = 4e-25

>TREMBL:CEM04B2_4 gene: "M04B2.4"; *Caenorhabditis elegans* cosmid M04B2
 Length = 527

HSPs:

Score = 801 (120.2 bits), Expect = 9.2e-80, P = 9.2e-80
 Identities = 171/433 (39%), Positives = 260/433 (60%)

```

Query:   61 PEHSDVVIIVGGVGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS 120
          P +++VI+GGG+ G S A+WLK+ R +V+VVE + ++++ST LS GGI QQFS
Sbjct:   91 PYRAEIVIIGGGLSGSSTAFWLKE-RFRDEDFKVVVVENNDVFTKSSTMLSTGGITQQFS 149

Query:  121 LPENIQLSLFSASFLRNINEYLAVVDAPPLDLRFNPSGYLLA-SEKDAAAMESNVKVQR 179
          +PE + +SLF+ FLR+ E+L ++D+ D+ F P+GYL LA ++++ M S KVQ
Sbjct:  150 IPEFVMSLFTTEFLRHAGEHLRILDSEQPDINFPTGYLRLAKTDEEVEEMMRSAWKVQI 209

Query:  180 QEGAKVSLMSPDQLRNKFPWINTGVALASYGMEDEGWFDWPCLLQGLRRKVQSLGVLFC 239
          + GAKV L+S D+L ++P++N + V LAS G+E+EG D W LL +R K +LGV +
Sbjct:  210 ERGAKVQLLSKDELTKRYPYMNVDVLLASLGVENEGTIDTWQLLSAIREKNITLGVQYV 269

Query:  240 QGEVTRFVSSSQRM-----LTDDKAVVLKRIHEVHVKMDRS-LEYQPVECAIVI 288
          +GEV F R T D+ + +RI V V+ + +P+ +++
Sbjct:  270 KGEVEGFQFERHRASSEVHAFGDDATADENKLAQRISGVLRPQMNDASARPIRAHLIV 329

Query:  289 NAAGAWSAQIAALAGVGEGPPGTLQGTKLPEVPRKRYVYVWHCPQGPGLTLPVADTS-G 347
          NAAG W+ Q+A +AG+G+G G L +P++PRKR V+V P P + P + D S G
Sbjct:  330 NAAGPWAGQVAKMAGIGKT-GLL-AVPVPIQPRKRDVEVIFAPDVPD-DLPFIIDPSTG 386

Query:  348 AYFRREGLGSNYLGGSPTEQEEP--DPANLEVDHDFQDKVWPHLALRVPAFETLKVS 405
          + R+ G +L GR+P+++E+ D +NL+VD+D F K+WP L RVP F+T KV+S
Sbjct:  387 VFCRQTDSGQTFVGRTPSKEADAKRDHNSLDVDYDDEYQKIWPVLVDRVPFGFQTAKVKS 446

```


Query: 406 AWAGYYDNTFDQNGVVGPHPLVVNMYFATGFSGHGLQQAPGIGRAVAEMVLKGRFQTID 465
 AW+GY D NTFD V+G HPL N++ GF G+ + RA AE + G + ++
 Sbjct: 447 AWSGYQDINTFDAPVIGEHPLYTNLHMMCGFGERGVMHSMMAARAYAEIRIFDGAYINVN 506

Query: 466 LSPFLFTRFYLGEKIQE 482
 L F R + I E
 Sbjct: 507 LRKFDMMRRIVKMDPITE 523

Pedant information for DKFZphutel_20b19, frame 3

Report for DKFZphutel_20b19.3

[LENGTH] 486
 [MW] 53811.85
 [pI] 7.66
 [HOMOL] TREMBL:CEM04B2_4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2 1e-78

[FUNCAT] c energy conversion [H. influenzae, HI0499] 8e-05
 [BLOCKS] BL00677A D-amino acid oxidases proteins
 [BLOCKS] BL00623A GMC oxidoreductases proteins
 [BLOCKS] BL01304A
 [EC] 1.5.99.2 Dimethylglycine dehydrogenase 2e-07
 [PIRKW] flavoprotein 2e-07
 [PIRKW] oxidoreductase 2e-07
 [PROSITE] MYRISTYL 12
 [PROSITE] CK2_PHOSPHO_SITE 5
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 6
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 7.00 %

SEQ MIRRVLPHGMGRGLLTRRPGTRRGFSLDWDGKVSEIKKKIKSILPGRSCDLLQDTSHP
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 PRD cccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccc
 MEM

SEQ PEHSDVVIIVGGVGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 PRD cccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccc
 MEMMM

SEQ LPENIQLSLFSASFRLRNINEYLAVVDAPPLDLRFNPSGYLLLAASEKDAAAMESNVKVRQ
 SEG
 PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccc
 MEM

SEQ EGAKVSLMSPDQLRNKFPWINTEGVALASYGMEDEGWFDPCWLLQGLRRKQSLGVLFCQ
 SEG
 PRD cccccccccchhhhhhhcc
 MEM

SEQ GEVTRFVSSQRMLTTDDKAVVLKRIHEVHVKMDRSLEYQVVECAIVINAAGAWSAQIAA
 SEG
 PRD cccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccc
 MEM

SEQ LAGVGEGPPGTQGTCLPVEPRKRYVYVWHCPQGPGLTPLVADTSGAYFRREGLSNYL
 SEG
 PRD hhcc
 MEM

SEQ GGRSPTEQEEPDPANLEVDHDFQDKVWPHLALRVPAFETLKVQSAWAGYYDNTFDQNG
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccc
 MEM

SEQ VVGPHPLVVNMYFATGFSGHGLQQAPGIGRAVAEMVLKGRFQTIDLSPLFTRFYLGEKI
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccc
 MEM

SEQ QENNII
 SEG
 PRD ccccc
 MEM

Prosites for DKFZphut1_20b19.3

PS00002	438->442	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	87->90	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	364->368	CK2_PHOSPHO_SITE	PDOC00006
PS00006	366->370	CK2_PHOSPHO_SITE	PDOC00006
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00008	354->360	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_20b19.3)

DKFZphutel_20g21

group: signal transduction

DKFZphutel_20g21 encodes a novel 861 amino acid protein with partial similarity to human ras inhibitor and other ras inhibitor proteins.

Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. The novel protein seems to be a new ras inhibitor protein.

The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

Ras inhibitor

additional 1188 Bp at 5' and 1107 at 3' end in comparison to I22483

Sequenced by AGOWA

Locus: unknown

Insert length: 4137 bp

Poly A stretch at pos. 4116, no polyadenylation signal found

```
1 GGGAGAACTG AAACAGGAGA TGGTGCGGAC AGATGTCAAC CTGGAAAATG
51 GCCTGGAACC CGCTGAAACC CACAGCATGG TAAGACACAA GGATGGTGGC
101 TATTCCGAGG AAGAGGACGT GAAGACCTGT GCCCGGGACT CAGGCTATGA
151 CAGCCTCTCC AACAGGCTCA GCATCTTGGA CCGGCTCCTC CACCCCCACC
201 CCATATGGCT GCAGCTGAGT CTGAGTGAGG AGGAGGCAGC AGAGTCTCTG
251 CAGGCCCAGC CTCCGGGGAT CTTCTGGTGT CATAAATCTA CCAAGATGCA
301 GAAGAAAGTC CTCTCCCTCC GCCTGCCCTG TGAATTTGGG GCCCCACTCA
351 AGGAATTGTC CATAAAGGAA AGCACATACA CCTTTCCCTT GGAAGGCTCA
401 GGAATCAGTT TCGCAGATTT ATTCCGGCTC ATTGCTTTCT ACTGCATCAG
451 CAGGGATGTT CTACCATTTA CCTTGAAGTT GCCTTATGCC ATTTCAACAG
501 CCAAGTCGGA GGCTCAGCTT GAAGAAGTGG CCCAGATGGG ACTAAATTTC
551 TGGAGCTCCC CAGCTGACAG CAAACCCCGG AACCTTCCAC CTCCCCATAG
601 GCCTCTTTCC TCCGACGGTG TCTGTCTGTC CTCCCTGCGT CAGCTCTGCC
651 TTATAAATGG AGTGCATTCT ATCAAAACCA GGACGCTTTC AGAGCTGGAG
701 TGCAGCCAGA CCAACGGGGC CCTGTGCTTT ATTAATCCCC TTTTCTTGAA
751 AGTGACACAG CAGGACCTCA GTGGAGGCCT GAAACGGCCG AGCACAAGGA
801 CTCCCAACGC GAATGGCAGC GAGCGGACTC GGTCCCCCCC ACCCAGGCCC
851 CCGCCACCCG CTATTAATAG TCTCCACACA AGCCCTCGGC TGGCCAGGAC
901 TGAACCCAGC ACGAGCATGC CAGAAACAGT CAACCATAAC AAACATGGGA
951 ACGTAGCTCT GCCTGGAACG AAACCAACTC CCATCCCTCC ACCCCGGCTG
1001 AAGAAGCAGG CTTCTTTTCT GGAAGCAGAG GGCGGTGCAA AGACCTTGAG
1051 CGGCGGCCCG CCGGCGCGAG GCCCGGAGCT GGAGCTGGGC ACAGCTGGCA
1101 GCCCAGGTGG GGCCCCGCCT GAGGCCGCCC CGGGGGATTG CACAAGGGCC
1151 CCGCCGCCCA GCTCTGAATC ACGGCCCCCG TGCCATGGAG GCCCGCAGCG
1201 GCTGAGCCAG ATGAGCATTT CTACTTCTCT CTCCGACTCG CTGGAGTTCTG
1251 ACCGGAGCAT GCCTCTGTTT GGCTACGAGG CGGACACCAA CAGCAGCCTG
1301 GAGGACTACG AGGGGGAAAG TGACCAAGAG ACCATGGCGC CCCCCATCAA
1351 GTCCAAAAGG AAAAGGAGCA GCTCCTTCGT GCTGCCCAAG CTCGTCAAGT
1401 CCCAGCTGCA GAAGGTGAGC GGGGTGTTCA GCTCCTTCAT GACCCCGGAG
1451 AAGCGGATGG TCCGCAGGAT CGCCGAGCTT TCCCGGGACA AATGCACCTA
1501 CTTGGGGTGC TTAGTGCAGG ACTACGTGAG CTTCTGCGAG GAGAACAAGG
1551 AGTGCCACGT GTCCAGCACC GACATGTGTC AGACCATCCG GCAGTTCATG
1601 ACCCAGGTCA AGAATATTT GTCTCAGAGC TCGGAGCTGG ACCCCCCCAT
1651 CGAGTCGCTG ATCCCTGAAG ACCAAATAGA TGTGGTGCTG GAAAAAGCCA
1701 TGCACAAGTG CATCTTGAAG CCCCTCAAGG GGCATGTGGA GGCCATGCTG
1751 AAGGACTTTC ACATGGCCGA TGGCTCATGG AAGCAACTCA AGGAGAACCT
1801 GCAGCTTGTG CGGCAGAGGA ATCCGCAGGA GCTGGGGGTC TTCGCCCCGA
1851 CCCCTGATTT TGTGGATGTG GAGAAAATCA AAGTCAAGTT CATGACCATG
1901 CAGAAGATGT ATTCGCCGGA AAAGAAGGTC ATGCTGCTGC TCGGGGTCTG
1951 CAAGCTCATT TACACGGTCA TGGAGAACAA CTCAGGGAGG ATGTATGGCG
2001 CTGATGACTT CTTGCCAGTC CTGACCTATG TCATAGCCCA GTGTGACATG
2051 CTTGAATTGG AACTGAAAT CGAGTACATG ATGGAGCTCC TAGACCCATC
2101 GCTGTTACAT GGAGAAGGAG GCTATTACTT GACAAGCGCA TATGGAGCAC
2151 TTTTCTGTAT AAAGAATTTT CAAGAAGAAC AAGCAGCGCG ACTGCTCAGC
2201 TCAGAAACCA GAGACACCTT GAGGCAGTGG CACAACCGGA GAACACCAA
2251 CCGGACCATC CCCTCTGTGG ACGACTTCCA GAATTACCTC CGAGTTGCAT
2301 TTCAGGAGGT CAACAGTGGT TGCACAGGAA AGACCTCCTT TGTGAGACCT
2351 TACATCACCA CTGAGGATGT GTGTGAGATC TGCGCTGAGA AGTTCAAGGT
2401 GGGGGACCCCT GAGGAGTACA GCCTCTTTCT CTTGCTTGAC GAGACATGGC
2451 AGCAGCTGGC AGAGGACACT TACCCTCAA AAATCAAGGC GGAGCTGCAC
```



```

2501 AGCCGACCAC AGCCCCACAT CTTCCACTTT GTCTACAAAC GCATCAAGAA
2551 CGATCCTTAT GGCATCATTT TCCAGAACGG GGAAGAAGAC CTCACCACCT
2601 CCTAGAAGAC AGGCGGGACT TCCCAGTGGT GCATCCAAAG GGGAGCTGGA
2651 AGCCTTGCCCT TCCCGCTTCT ACATGCTTGA GCTTGAAAAG CAGTCACCTC
2701 CTCGGGGACC CCTCAGTGTA GTGACTAAGC CATCCACAGG CCAACTCGGC
2751 CAAGGGCAAC TTTAGCCACG CAAGGTAGCT GAGGTTTGTG AAACAGTAGG
2801 ATTCTCTTTT GGCAATGGAG AATTGCATCT GATGGTTCAA GTGTCTCTGAG
2851 ATTGTTTGCT ACCTACCCCC AGTCAGGTTT TAGGTTGGCT TACAGGTATG
2901 TATATGTGCA GAAGAAACAC TTAAGATACA AGTTCTTTTG AATTCAACAG
2951 CAGATGCTTG CGATGCAGTG CGTCAGGTGA TTCTCACTCC TGTGGATGGC
3001 TTCATCCCTG CCTTCCTTCC TTTCTTTTTC CTTTTTTTTT TTTTTTTTTT
3051 TTTTACAAA GAGCCTTCAT GTTTTTATAT ATTTATAGA AATTTTATA
3101 GCAGTTGCAG GTAAACTGTC AGGATTGGTT TAAAAATATT TTTGTAACCT
3151 TAAATATTTC TATAATTATG CATGTGATT TAAACATTTAA TATTCAAAAA
3201 TAAATCTCTT GCTGGATTG AGAGTATTGC ATTTTAAAG TCTCTCTTCT
3251 GTAACCTGGT GTTTTGGCAA CTTTGTGGGG AGAGACTGCT GGATTCTTAA
3301 AAGCAACGTA TTCCTGACAC TGGCCACAGA ATGCCTTTGG AAATCGGATG
3351 TACTGTTCTC TTGTTACGCT TTAGTGGTGT TTTGCTGTTT TGTTTTTAA
3401 ACAATATGAT CTGAGAATAA GGAGAGAAAT GAATGTAGAG AGAGGTAGAG
3451 AGAGAAATAT GAACTCTAAC AAAGGACTGA GGAGTGCAGT CTGCTGGTTC
3501 AGGCTCTTCA AAAGATGTAG AAAAGAGAT AGAAGGAACC ACCTATGCTT
3551 AAAATACTGT AAATATGCAG TGAGGTTTGG CAAAATCTAT TCCATGTGTG
3601 ATTTGCTTGT AGAAACAATT TTGAAAGCCC CTTGAGGAAA ATAAAAATCA
3651 AGAAGAACAC TTTTCTCCCT TTCCATACA AATTAATACT TAACAGCATC
3701 AAATTATTGG GACCAGAAAC CAAGTAATGT ATAATGTGGC TTTTGTGAG
3751 TTAATAAAGA TGCTATATAA TGGAGAAGAA TTTGAAATG CACAAAAAAA
3801 TCAATCTACA TTATCAGAAC CTGCAGTGAA ATTAACCTTA TGTTAAATAA
3851 AACCAGTTTG CAGGTGCACA AACTATGAGG GTCTTGATC CACGTAACAC
3901 AGGTAGTTAC AAAAACATGT TATTGTACTG TGTAAGATG CATAGTCATC
3951 TCATTTGGTT GGCTTTGTAC CTTGTACCTT TTTTAGCCTT GGCTTTTGT
4001 GAACTAGAAC CTCAGCACA TACTGTGTTG TACTTTTGT AATGATTTT
4051 TAAATGGAAT TTTGCACATA ATACATTGTA ATACTGTATG ATAATCATGT
4101 GTGAAAATAA TTTTGAAT AAAAATAA AAAAAA

```

BLAST Results

Entry I22483 from database EMBL:
Sequence 15 from patent US 5527896.
Length = 1829
Plus Strand HSPs:
Score = 9097 (1364.9 bits), Expect = 0.0, P = 0.0
Identities = 1821/1823 (99%), Positives = 1821/1823 (99%),

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2602 bp; peptide length: 861
Category: known protein
Classification: Cell signaling/communication

```

1 MVRTDVNLEN GLEPAETHSM VRHKDGGYSE EEDVKTCARD SGYDSLSNRL
51 SILDRLLHTH PIWLQLSLSE EEAEEVLQAO PPGIFLVHKS TKMQKKVLSL
101 RLPCEFGAPL KEFAIKESTY TFSLEGSGIS FADLFRLIAF YCISRDLVLF
151 TLKLPYAIST AKSEAQLEEL AQMGLNFWSS PAOSKPPNLP PPHRPLSSDG
201 VCPASLRQLC LINGVHSIKT RTPSELECSQ TNGALCFINP LFLKVHSQDL
251 SGLKRPSTR TPNANGTERT RSPPPRPPPP AINSLHTSPR LARTETQTSM
301 PETVNNHKGH NVALPGTKPT PIPFRLKKQ ASFLEAEGGA KTLSGGRPGA
351 GPELELGTAG SPGGAPPEAA PGDCTRAPPP SSESPPCHG GRQLSDMSI
401 STSSSDSLEF DRSMPLFGYE ADTNSSLEDY EGESDQETMA PPIKSKKKRS
451 SSFVLPKLVK SQLQKVSGVF SSMTPPEKRM VRRIAELSRD KCTYFGCLVQ
501 DVVSFLQENK ECHVSSDML QTIQFMTQV KNYLSQSSSEL DPPIESLIPE
551 DQIDVLEKA MHKILKPLK GHVEAMLKDF HMADGSKQL KENLQLVRQR
601 NPQELGVFAP TPDEFVDEKI KVKFMTMQKM YSPEKKVMLL LRVCKLIYTV
651 MENNSGRMYG ADDELPLVLT VIAQCDMLEL DTEIYMMEL LDPSLLHGEG
701 GYVLTSA YCA LSLIKNFQEE QAARLLSSET RDTLRQWHRK RTNRTIPSV
751 DDFQNYLRVA FQEVNSGCTG KTLVLRPYIT TEDVCQICAE KFKVGDPPEY
801 SLFLFVDETQ QQLAEDTYPQ KIKAEHLSRP QPHIFHFVYK RIKNDPYGII
851 FQNGEEDLTT S

```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20g21, frame 2

TREMBL:RNU80076_1 product: "RIN1"; Rattus norvegicus RIN1 mRNA, complete cds., N = 3, Score = 606, P = 6.8e-97

PIR:A38637 Ras interactor RIN1 - human, N = 3, Score = 587, P = 1.9e-92

TREMBL:HSRASINL_1 product: "ras inhibitor"; Human ras inhibitor mRNA, 3' end., N = 2, Score = 592, P = 9.8e-61

SWISSPROT:RIN1_HUMAN RAS INTERACTION/INTERFERENCE PROTEIN 1 (RAS INHIBITOR JC99) (FRAGMENT)., N = 2, Score = 587, P = 4.1e-60

PIR:B38637 Ras inhibitor (clone JC265) - human (fragment), N = 1, Score = 2446, P = 4.6e-254

>PIR:B38637 Ras inhibitor (clone JC265) - human (fragment)
Length = 471

HSPs:

Score = 2446 (367.0 bits), Expect = 4.6e-254, P = 4.6e-254
Identities = 471/471 (100%), Positives = 471/471 (100%)

```

Query:   391 GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 450
          GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS
Sbjct:   1 GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 60

Query:   451 SSFVLPKLVKSQLQKVSGVFSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 510
          SSFVLPKLVKSQLQKVSGVFSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK
Sbjct:   61 SSFVLPKLVKSQLQKVSGVFSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 120

Query:   511 ECHVSSTDMQLTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 570
          ECHVSSTDMQLTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK
Sbjct:   121 ECHVSSTDMQLTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 180

Query:   571 GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 630
          GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM
Sbjct:   181 GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 240

Query:   631 YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTIVIAQCMDLELDTEIEYMMEL 690
          YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTIVIAQCMDLELDTEIEYMMEL
Sbjct:   241 YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTIVIAQCMDLELDTEIEYMMEL 300

Query:   691 LDPSLLHGEAGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV 750
          LDPSLLHGEAGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV
Sbjct:   301 LDPSLLHGEAGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV 360

Query:   751 DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 810
          DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW
Sbjct:   361 DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 420

Query:   811 QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 861
          QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS
Sbjct:   421 QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 471

```

Pedant information for DKFZphutel_20g21, frame 2

Report for DKFZphutel_20g21.2

```

[LENGTH]      861
[MW]           96380.26
[pI]           6.15
[HOMOL]        PIR:B38637 Ras inhibitor (clone JC265) - human (fragment) 0.0
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YML097c] 3e-10
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
3e-10
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YML097c] 3e-10
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
3e-10
[PIRKW]        alternative splicing 3e-59
[SUPFAM]       Ras interactor RIN1 3e-59

```



```

SEQ      MVRTDVLNLENGLEPAETHSMVRHKDGGYSEEDVKTCCARDSGYSLSNRLSILDRLLHTT
SEG      .....
PRD      cccccceccccccccccccceccccccccccccceccccccccchhhhhhhhhhhhhhhhh

SEQ      PIWLQLSLSEEEAAEVLOAQPPIFLVHKSTKMOKKVLSLRLPCEFGAPLKEFAIESTY
SEG      . . . xxxxxxxxxxxxxxxxxxx . . .
PRD      hhhhhhhhhhhhhhhhhhhhhhhcccccceeeechhhhhhhhhhhhhccccccccccccceeececc

SEQ      TFSLESGSISFADLFRLIAFYCISRDLVLPFTLKLPHYAISTAKSEAQLEELAQMGLNFWSS
SEG      .....
PRD      cececcccchhhhhhhhhhhhhhhccceeececcccchhhhhhhhhhhhhhhhhhhhhhhcccccc

SEQ      PADSKPPNLPPPHRLPSDGVCPASLRQLCLINGVHSIKTRTPSELECSQTNGALCFINP
SEG      . . . xxxxxxxxxxx . . .
PRD      cccccccccccccccccccccccccchhhhhhhccccccccccccccccccccccccceeecc

SEQ      LFLKVHSQDLSGGLKRPSTRTPNANGTERTRSPPRPPPPAINSLHTSPRLARTETQTSM
SEG      . . . xxxxxxxx . . .
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PETVNHKNHGNVALPGTKPTPIPPRLKKQASFLAEGGAKTLSGGRPGAGPELELGTAG
SEG      . . . xxxxxxxxxxx . . .
PRD      eeececcccccccccccccccccccchhhhhhhhhhhhhccccccccccccccccceeececc

SEQ      SPGGAPPEAAPGDCTRAPPPSSERPPCHGGRQLSDMSISTSSSDSLEFDRSMPLFGYE
SEG      xxxxxxxxxxx . . . xxxxxxxxxxxxxxxxxxxxxxx . . .
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccceeeccccccceee

SEQ      ADTNSSLEDYEGESDQETMAPPIKSKKKRSSSFVLPKLVKSQQLQKVGVSFSSFMTPEKRM
SEG      . . . xxxxxxxx . . .
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhcchhhh

SEQ      VRRIAELSRDKCTYFGCLVQDYVSFLQENKECHVSSSTDMLOQTIRQFMTQVKNYLSQSSEL
SEG      .....
PRD      hhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhc

SEQ      DPPIESLIPEDQIDVVLKAMHKCILKPLKGHEAMLKDFHMDAGSWGKQLKENLQVLRQR
SEG      .....
PRD      cccccccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhh

SEQ      NPQELGVFAPTPDFVDVEKIKVKFMTMQMYSPEKKVMLLLRVCKLIYTMENNSSGRMYG
SEG      .....
PRD      cccccccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhcccccc

SEQ      ADDFLPVLTYVIAQCDMLELDEIEYMMELLDPSSLHGEGGYLTSAYGALS LIKNFQEE
SEG      .....
PRD      cccccccceccccchhhhhhhhhhhhhhhccccccccceeeehhhhhhhhhhhhhhhhh

SEQ      QAARLLSSETRDTRLQWHKRRRTNRTIPSVDDFQNYLRVAFQEVNSGCTGKTLVRPYIT
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhccccccccceeececccccc

SEQ      TEDVCQICAEKFKVGDPEEYSLFLVDETWQQLAEDTYPQKIKAEHSRPPQPHIFHFVYK
SEG      .....
PRD      chhhhhhhhhheeeccccceeeehhhhhhhccccccccchhhhhhhhhhhccccceeeehhh

SEQ      RIKNDPYGIIFQNGEEDLTT
SEG      .....
PRD      hhccccceeecccccccccc

```

(No Pfam data available for DKFZphute1_20g21.2)

DKFZphutel_20h13

group: intracellular transport and trafficking

DKFZphutel_20h13 encodes a novel 955 amino acid protein with similarity to alpha-adaptins.

Adaptins are components of the adaptor complexes which link clathrin to receptors in coated vesicles. The alpha-adaptins, which are found exclusively in endocytic coated vesicles, separate into two bands on SDS gels, designated A and C. The novel protein is very similar to both alpha adaptin A and C. The novel protein is a new human alpha-adaptin.

The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

strong similarity to alpha-adaptins

complete cDNA, complete cds start at Bp 78, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3352 bp

Poly A stretch at pos. 3297, polyadenylation signal at pos. 3279

```

1 GCGCCCGGTC CCCGCTTGCC AGCCCCCGCT GCTCTGTGCC CTGTCCGGCC
51 AGGCCTGGAG CCGACACCAC CGCCATCATG CCGGCCGTGT CCAAGGGCGA
101 TGGGATGCGG GGGCTCGCGG TGTTTCATCTC CGACATCCGG AACTGTAAGA
151 GCAAAGAGGC GGAATTAAG AGAATCAACA AGGAAGTGGC CAACATCCGC
201 TCCAAGTTCA AAGGAGACAA AGCCTTGGAT GGCTACAGTA AGAAAAAATA
251 TGTGTGTAAA CTGCTTTTCA TCTTCTGCT TGGCCATGAC ATTGACTTTG
301 GGCACATGGA GGCTGTGAAT CTGTTGAGTT CCAATAAATA CACAGAGAAG
351 CAAATAGGTT ACCTGTTTCA TTCTGTGCTG GTGAACTCGA ACTCGGAGCT
401 GATCCGCGTC ATCAACAACG CCATCAAGAA TGACCTGGCC AGCCGCAACC
451 CCACCTTCAT GTGCCTGGCC CTGCACTGCA TCCCAACGT GGGCAGCCGG
501 GAGATGGGCG AGGCCTTTGC CGCTGACATC CCCCGCATCC TGGTGGCCGG
551 GGACAGCATG GACAGTGTCA AGCAGAGTGC GGCCCTGTGC CTCCTTCGAC
601 TGTACAAGGC CTCGCCTGAC CTGGTGCCCA TGGGCGAGTG GACGGCGCGT
651 GTGGTACACC TGCTCAATGA CCAGCACATG GGTGTGGTCA CGGCCGCCGT
701 CAGCCTCATC ACCTGTCTCT GCAAGAAGAA CCCAGATGAC TTCAAGACGT
751 GCGTCTCTCT GGCTGTGTCT CGCCTGAGCC GGATCGTCTC CTCTGCCTCC
801 ACCGACCTCC AGGACTACAC CTACTACTTC GTCCAGCAC CCTGGCTCTC
851 GGTGAAGTCT CTGCGGCTGC TGCAGTGCTA CCGGCTTCCA GAGGATGCGG
901 CTGTGAAGGG GCGGCTGGTG GAATGTCTGG AGACTGTGCT CAACAAGGCC
951 CAGGAGCCCC CCAATCCAA GAAGTGCAG CATTCCAACG CCAAGAACGC
1001 CATCCTCTTC GAGACCATCA GCCTCATCAT CCACTATGAC AGTGAGCCCA
1051 ACCTCCTGGT TCGGGCCTGC AACCAGCTGG GCCAGTTCTT GCAGACCCGG
1101 GAGACCAACC TGGCTACCTT GGCCCTGGAG AGCATGTGCA CGCTGGCCAG
1151 CTCGAGTTTC TCCATGAAG CCGTCAAGAC GCACATTGAC ACCGTCATCA
1201 ATGCCCTCAA GACGGAGCGG GACGTCAGCG TCGGCGAGCG GCGGCGCTGAC
1251 CTCCTCTACG CCATGTGTGA CCGGAGCAAT GCCAAGCAGA TCGTGTCCGA
1301 GATGCTGCGG TACCTGGAGA CGGCAGACTA CGCCATCCGC GAGGAGATCG
1351 TCCTGAAGGT GGCCATCCTG GCCGAGAAGT ACGCCGTGGA CTACAGCTGG
1401 TACGTGGACA CCATCCTCAA CCTCATCCGC ATTGCGGGCG ACTACGTGAG
1451 TGAGGAGGTG TGGTACCGTG TGCTACAGAT CGTCACCAAC CGTGATGACG
1501 TCCAGGGGTA TGCCGCCAAG ACCGTCTTTG AGGCGCTCCA GGCCCTGCC
1551 TGTACAGAGA ACATGGTGAA GGTGCGCGG TACATCCTTG GGGAGTTTGG
1601 GAACCTGATT GCTGGGGACC CCGCTCCAG CCCCCAGTG CAGTTCTCCC
1651 TGCTCCACTC CAAGTTCCAT CTGTGCAGCG TGGCCACGCG GCGGCTGCTG
1701 CTGTCCACCT ACATCAAGTT CATCAACCTC TTCCCGGAGA CCAAGGCCAC
1751 CATCCAGGGC GTCCTGCGGG CCGGCTCCCA GCTGCGCAAT GCTGACGTGG
1801 AGCTGCAGCA GCGAGCCGTG GAGTACCTCA CCCTCAGCTC AGTGGCCAGC
1851 ACCGACGTCC TGGCCACGGT GCTGGAGGAG ATGCCGCCCT TCCCGAGCG
1901 CGAGTCGTCC ATCCTGGCCA AGCTGAAACG CAAGAAGGGG CCAGGGGCGG
1951 GCAGCGCCCT GGACGATGGC CGGAGGGACC CCAGCAGCAA CGACATCAAC
2001 GGGGGCATGG AGCCCAACCC CAGCACTGTG TCGACGCCCT CGCCCTCCGC
2051 CGACCTCCTG GGGCTGCGGG CAGCCCTTCC CCGGCGAGCA CCCCCGGCTT
2101 CTGCAGGAGC AGGGAACCTT CTGGTGGACG TCTTCGATGG CCGGCGCCGC
2151 CAGCCAGGCC TGGGGCCAC CCCCAGGAG GCCTTCTCTA GCCCAGGTCC
2201 TGAGGACATC GGCCCTCCCA TTCCGGAAGC CGATGAGTTG CTGAATAAGT
2251 TTGTGTGTAA GAACAACGGG GTCCTGTTCC AGAACAGCT GCTGCAGATC
2301 GGAGTCAAGT CAGAGTTCCG ACAGAACCTG GGCCGATGT ATCTCTTCTA
2351 TGGCAACAAG ACCTCGGTGC AGTTCCAGAA TTTCTCACC ACTGTGGTTC
2401 ACCCGGGAGA CCTCCAGACT CAGCTGGCTG TGCAGACCAA GCGCGTGGCG
2451 GCGCAGGTGG ACGGCGGCGC GCAGGTGCAG CAGGTGCTCA ATATCGAGTG
2501 CCTGCGGGAC TTCTGACGCG CCCCCTGCT GTCCGTGCGC TTCCGTGACG
2551 GTGGCGCCCC CCAGGCCCTC ACCCTGAAGC TCCAGTGAC CATCAACAAG

```



```

2601 TTCTTCCAGC CCACCGAGAT GGCGGCCAG GATTCTTCC AGCGCTGGAA
2651 GCAGCTGAGC CTCCCTCAAC AGGAGGCGCA GAAATCTTC AAAGCCAACC
2701 ACCCCATGGA CGCAGAAGTT ACTAAGGCCA AGCTTCTGGG GTTTGGCTCT
2751 GCTCTCCTGG ACAATGTGGA CCCCACCCCT GAGAACTTCG TGGGGGGGGG
2801 GATCATCCAG ACTAAGCCC TGCAGGTGGG CTGTCTGCTT CGGCTGGAGC
2851 CCAATGCCCA GGCCCAAGATG TACCGGCTGA CCCTGCGCAC CAGCAAGGAG
2901 CCCGTCTCCC GTCACCTGTG TGAGCTGCTG GCACAGCAGT TCTGAGCCCT
2951 GGACTCTGCC CCGGGGGATG TGGCCGGCAC TGGGCAGCCC CTTGGACTGA
3001 GGCAGTTTTC GTGGATGGGG GACCTCCACT GGTGACAGAG AAGACACCAG
3051 GGTTTGGGGG ATGCCTGGGA CTTTCTCCG GCCTTTTGTA TTTTATTTT
3101 TGTTTCATCTG CTGCTGTTTA CATTCTGGGG GGTAGGGGG AGTCCCCCTC
3151 CCTCCCTTTC CCCCCAAGC ACAGAGGGGA GAGGGGCCAG GGAAGTGGAT
3201 GTCTCTCTCC CTCCACCCC ACCCTGTTGT AGCCCCCTCT ACCCCCTCCC
3251 CATCCAGGGG CTGTGTATTA TTGTGAGCGA ATAAACAGAG AGACGCTAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AA

```

BLAST Results

No BLAST result

Medline entries

89155572:
Cloning of cDNAs encoding two related 100-kD coated vesicle proteins (alpha-adaptins).

97431776:
Alpha-adaptin, a marker for endocytosis, is expressed in complex patterns during Drosophila development.

Peptide information for frame 3

ORF from 78 bp to 2942 bp; peptide length: 955
Category: strong similarity to known protein

```

1 MPAVSKGDGM RGLAVFISDI RNCKSKEAEI KRINKELANI RSKFKGDKAL
51 DGYSKKKYVC KLLFIFLLGH DIDFGHMEAV NLLSSNKYTE KQIGYLFISV
101 LVNSNSELI R LINNAIKNDL ASRNPFTMCL ALHCIANVGS REMGEAFAAD
151 IPRIILVAGDS MDSVKQSAAL CLLRLYKASP DLVPMGEWTA RVVHLLNDQH
201 MGVVTAAVSL ITCLCKKNPD DFKTCVSLAV SRLSRIVSSA STDLDQDYTY
251 FVPAPWLSVK LLRLLCYPP PEDAAVKGRL VECLETVLNK AQEPPKSKKV
301 QHSNAKNAIL FETISLIHY DSEPNLLVRA CNQLGQFLOH RETNRLYLAL
351 ESMCTLASSE FSHEAVKTHI DTVINALKTE RDVSVRQRAA DLYIAMCDRS
401 NAKQIVSEML RYLETADYAI REEIVLKVAI LAEKYAVDYS WYVDTILNLI
451 RIAGDYVSEE VWYRVLQIVT NRDDVQGYAA KTVFEALQAP ACHENMVKVG
501 GYILGEFGNL IAGDPRSSPP VQFSLHLSKF HLCSVATRAL LLSTYIKFIN
551 LFPETKATIQ GVLRAQSRLR NADVELQORA VEYLTLSVA STDVLATVLE
601 EMPFPFERES SILAKLKRKK GPGAGSALDD GRDPSSNDI NGGMEPTPST
651 VSTPSPSADL LGLRAAPPFA APPASAGAGN LLVDVFDGPA AQPSLGPTPE
701 EAFSLPGPED IGPIPEADE LLNKFVCKNN GVLFFENQLLQ IGVKSEFRON
751 LGRMYLFYGN KTSVQFQNF SPTVVHPGDLQ TQLAVQTKRV AAQVDGGAQV
801 QQVLNIECLR DFLTPPLLSV RFRYGGAPQA LTLKLPVTIN KFFQPTEMAA
851 QDFFQRWKQL SLPOQEAQKI FKANHPMDAE VTKAKLLGFG SALLDNVDPN
901 PENFVGAGII QTKALQVGCL LRLEPNAQAA MYRLTLRTSK EPVSRHLCEL
951 LAQQF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20h13, frame 3

PIR:B30111 alpha-adaptin C - mouse, N = 1, Score = 3990, P = 0

PIR:S11276 alpha-adaptin c - rat, N = 1, Score = 3987, P = 0

SWISSPROT:ADAC RAT ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT)., N = 1, Score = 3982, P = 0

SWISSPROT:ADAC_MOUSE ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX
2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA
MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT).. N = 1, Score =
3976, P = 0

TREMBL:AB020706_1 gene: "KIAA0899"; product: "KIAA0899 protein"; Homo
sapiens mRNA for KIAA0899 protein, partial cds., N = 1, Score = 3932, P
= 0

>PIR:B30111 alpha-adaptin C - mouse
Length = 938

HSPs:

Score = 3990 (598.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 787/955 (82%), Positives = 858/955 (89%)

```

Query:      1 MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60
            MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
Sbjct:      1 MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60

Query:     61 KLLFIFLLGHIDIDFGHMEAVNLLSSNRYTEKQIGYLFISVLVNSNELIRLINNAIKNDL 120
            KLLFIFLLGHIDIDFGHMEAVNLLSSNRYTEKQIGYLFISVLVNSNELIRLINNAIKNDL
Sbjct:     61 KLLFIFLLGHIDIDFGHMEAVNLLSSNRYTEKQIGYLFISVLVNSNELIRLINNAIKNDL 120

Query:    121 ASRNPTFMCLALHCIANVGSREMGEAFAADIPRILVAGDSMOSVKQSAALCLLRLYKASP 180
            ASRNPTFMCLALHCIANVGSREM EAF +IP+ILVAGD+MOSVKQSAALCLLRLY+ SP
Sbjct:    121 ASRNPTFMGLALHCIANVGSREMAEAFAGEIPKILVAGDTMOSVKQSAALCLLRLYRTSP 180

Query:    181 DLVPMGEWTARVVHLLNDQHMGGVTAASVSLITCLCKKNPDDFKTCVSLAVSRLSRIVSSA 240
            DLVPMG+WT+RVVHLLNDQH+GVVTA SLIT L +KNP++FKT VSLAVSRLSRIV+SA
Sbjct:    181 DLVPMGDWTSRVVHLLNDQHLGVVTAATSLITTLAQKNPEEFKTSVSLAVSRLSRIVTSA 240

Query:    241 STDLDQDYTYFYFVPAPWLSVKLLRLQCYPPEDAAVKGRLEVECTETVLNKAQEPKSKKV 300
            STDLDQDYTYFYFVPAPWLSVKLLRLQCYPPE D AV+GRL ECLET+LNKAQEPKSKKV
Sbjct:    241 STDLDQDYTYFYFVPAPWLSVKLLRLQCYPPE-DPAVRGRLTECLETILNKAQEPKSKKV 299

Query:    301 QHSNAKNAILFETISLIHYDSEPNLLVRACNQLGQFLQHRETNRLRYLALESMTCLASSE 360
            QHSNAKNA+LFE ISLIH+DSEPNLLVRACNQLGQFLQHRETNRLRYLALESMTCLASSE
Sbjct:    300 QHSNAKNAVLFEAISLIHHDSEPNLLVRACNQLGQFLQHRETNRLRYLALESMTCLASSE 359

Query:    361 FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAQIVSEMLRYLETADYAI 420
            FSHEAVKTHI+TVINALKTERDVSVRQRA DLYAMCDRSNA+QIV+EML YLETADY+I
Sbjct:    360 FSHEAVKTHIETVINALKTERDVSVRQRAVDLLYAMCDRSNAQIVAEMLSYLETADYSI 419

Query:    421 REEIVLKVAILAEKYAVDYSWYVDITILNLIRIAGDYVSEEVWYRVLQIVTNRDDVQGYAA 480
            REEIVLKVAILAEKYAVDY+WYVDITILNLIRIAGDYVSEEVWYRV+QIV NRDDVQGYAA
Sbjct:    420 REEIVLKVAILAEKYAVDYTWYVDITILNLIRIAGDYVSEEVWYRVQIVINRDDVQGYAA 479

Query:    481 KTVFEALQAPACHENMKVGGYILGEFGNLIAGDPRSSPPVQFSLHSHKFLHCSVATRAL 540
            KTVFEALQAPACHEN+VKVGGYILGEFGNLIAGDPRSSP +QF+LLHSHKFLHCSV TRAL
Sbjct:    480 KTVFEALQAPACHENLVKGGYILGEFGNLIAGDPRSSPLIQFNLHSHKFLHCSVPTRAL 539

Query:    541 LLSTYIKFINLFPETKATIQGVLRAGSQLRNADVELQQRAVEYLTLSVASTDVLATVLE 600
            LLSTYIKF+NLFP KATIQ VLR+ SQL+NADVELQQRAVEYL LS+VASTD+LATVLE
Sbjct:    540 LLSTYIKFVNLFPETKATIQDVLRSDSQLKNADVELQQRAVEYLRSTVASTDILATVLE 599

Query:    601 EMPFFPERESSILAKLRKKGPAGSALDDGRDPSSNDINGGMEPTF---STVSTPSPS 657
            EMPFFPERESSILAKL+KKGP + L++ +R+ S D+NGG EP P S STPSPS
Sbjct:    600 EMPFFPERESSILAKLKKKGPSTVTDL EETKRERSI-DVNGGPEPVPASTSAASTPSPS 658

Query:    658 ADLLGLRAAPP-PAAPPASAGAGNLLVDVFDGPAAPSLGPTPEEAFSLPGPEDIGPPIP 716
            ADLLGL A PP P PP S+G G LLVDVF A+ ++ P L+PG ED
Sbjct:    659 ADLLGLGAVPPAPTGPSSSGGG-LLVDVFSAS--AVAP-----LAPGSEDN----- 704

Query:    717 EADELLNKFCVCKNNGVLFENQLLQIGVKSEFRQNLGRMYLFYGNKTSVQFQNFSPTVVHP 776
            +FVCKNNGVLFENQLLQIG+KSEFRQNLGRM++FYGNKTS QF NF+PT++
Sbjct:    705 -----FARFVCKNNGVLFENQLLQIGLKSEFRQNLGRMFIFYGNKTSQFLNFTPTLICA 759

Query:    777 GDLOTQLAVQTKRVAQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 836
            DLQT L +QTK V VDGGAQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
Sbjct:    760 DDLOTNLNLQTKPVDPTVDGGAQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 819

Query:    837 VTINKFFQPTEMAAQDFFQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 896
            +T+KFFQPTEMA+QDFFQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
Sbjct:    820 ITINKFFQPTEMASQDFFQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 879

Query:    897 VDPNPENFVGAGIIQTKALQVGCLLRLLEPNAQAQMYRLTLRTSKPEVSRHLCELLAQGF 955
            VDPNP NFGAGII TK Q+GCLLRLEPN QAQMYRLTLRTSK+ VS+ LCELL++QF

```


Prosites for DKFZphute1_20h13.3

(No Pfam data available for DKFZphute1_20h13.3)

DKFZphutel_20ml1

group: cell cycle

DKFZphutel_20ml1 encodes a novel 225 amino acid protein with similarity to yeast sds22 and protein phosphatase-1 regulatory subunits.

sds22 is a regulatory polypeptide of protein phosphatase-1 that is required for the completion of mitosis in both fission and budding yeast. The novel protein seems to be a new regulator protein for protein phosphatase-1.

The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

similarity to suppressor protein sds22

complete cDNA, complete cds, EST hits
localisation? only a part of the STS matches

Sequenced by AGOWA

Locus: /map="17"?

Insert length: 5822 bp

Poly A stretch at pos. 5803, polyadenylation signal at pos. 5786

```

1 GGGCGCTTGG TTCCCCAGCA ACCGGGAGAC GCGTCTGCTG CGTGGAAACCG
51 CCGAGTTCCC AGCGCTTGAG AAGGAAAATT CTGGATCTGT TATCTGTGAG
101 GAGGCCACTC CGTTGACAGT TGTGTAAAC TCTGCTGCTT TCCCCAGCTC
151 CAACCTCTCT GGTCTTCAAC AACACTATCA TCAGGGAAAA CGTGGGGGAA
201 GATGAACCA GCGTGCAACT CGATGGAGCC GAGGGTGATG GACGATGACA
251 TGCTCAAGCT GGGCGTCGGG GACGAGGGCC CCCAGGAGGA GGCCGGGCAG
301 CTGGCCAAAG AGGAGGGCAT CCTCTTCAAG GATGTCCTGT CCCTGCAGCT
351 GGACTTTCGG AACATCCTCC GCATAGACAA CCTCTGGCAG TTTGAGAACT
401 TGAGGAAGCT GCAGCTGGAC AATAACATCA TTGAGAAGAT CGAGGGCCTG
451 GAGAACCCTG CACACCTGGT CTGGCTGGAT CTGTCTTTCA ACAACATTGA
501 GACCATCGAG GGGCTGGACA CACTGGTGAA CCTGGAGGAC CTGAGCTTGT
551 TCAACAACCG GATCTCCAAG ATCGACTCCC TGGACGCCCT CGTCAAGCTG
601 CAGGTGTTGT CGCTGGGCAA CAACCGGATT GACAACATGA TGAACATCAT
651 CTACCTCCGG CGGTTCAGT GCCTGCGGAC GCTCAGCCTC TCTAGGAACC
701 CTATCTCTGA GGCAGAGGAT TACAAGATGT TCATCTGTGC CTACCTTCCT
751 GACCTCATGT ACCTGGACTA CCGGCGCATT GATGACCACA CAGCAAGTGT
801 CTCCCTCTCA GTCTCCAGC CCTGTGAGAC AGATTCTCTA AGCCCCAGG
851 TTTCTTGGAA AAGGGGCATT GAAGAGTAGC TTCCCTGACC CACAACATAGG
901 AGAGAAAGGG CAGCTCCCTC TTCCTAATCC CTTTACCTGA CTCTGTGAGA
951 GTGATTCCAG CAGCACCCCT GTAAGTACTG TTTTGTGTGC GTTCCCAGGG
1001 GCCAGGCCCT TTCCACACAC TGTCCAGGG CCACCTCACA GCCATCCTGC
1051 ACTGTCTAGT TTTCCAGATG AAGAAGCTGA GGAGGGCTGG GAGCAGTGGC
1101 TCACGCCCTG AATCCCAGCA CTTTGAGAGG CTGAGGCGGG AGGATCGCTT
1151 GAGCCAAAGG GTTCAAGACC AGCCTGGGCA ACATAGGGAG ACCCCATCTC
1201 TACAGAAACT ACCAAAATTA GCCAGGTGTG GTGGCACACA CCAGTAATCC
1251 TGGCTACTCA CAAGGCCGAG GTAGAAGAA CGCTTGAGAC TAGGAGTTTG
1301 AGGCTGCAGT GAACTAAGAA GATGCCATTG CACTCCAGCC TGGGCAACAG
1351 AGTGAAAAAA TTAATAAATT AGAAAAGAAA AGAAGTTGAG GAGGCCCAAG
1401 GAGGGCAAGC AGCCAGGATC ACTGGCTCAA GGCCAAGCCA GGATTCAACC
1451 TAAGTTGGTG TCATCCCAGG AGCAATATTA ACAGCTGAGC TCCAGAGGGA
1501 ACCAGGCCAT CAGAGGCTCA GGCCTGGCTC TCAGGGGCAG AGTCAGGGCT
1551 GGAGGTAGAG ACCTGAGTGT CATCTGAGGA TTGCCAATTG GCAGTAGTTG
1601 AAGCCATGGT ACAGGTGGGA TCACCTGGGG CACATGGAGT GAGCTGGGGG
1651 ACGGGGACTA AGTTCTAGAG GTGCCAGCAT TCCTGGCCAG GTACAGGGGG
1701 ATGAGCCAGT GCGGTGGAGA GAGCCAAGGG CCAGACCCTC GTGACCAGCC
1751 CTATGGCCTC ACTTACCTC TGTCTGTGTG TCCTCCTTCC CTAAGAGAGG
1801 GCCAGAAGGC CTGCTGAGGG CTGTTGGGAG TGAGAGAGCA AGTCCTCTGT
1851 GGAGAACACC CAGTCTGGGG CGAGGGGAGC GCTCCATTGC TGTGGCTCCT
1901 GCCCTGGAGA TGGCCCCGGG AAGCCAGGCC TGCCACGCTG CCTCCGCTC
1951 CTCCTGGTCT TTCCCTGATT TCCCTGCGCT CACAAAAACC TGGTGAGGGT
2001 CATCAGGAGA TGGGCATTCT CATCCACGAG ACCTCATGGC TTTCACAGCC
2051 TTCTATGAGG CCCCCTGTGA ACACCCCTGC CCATGCGCGG GAGGCTGCAG
2101 CATGGCAGAG CGGCATGGC AGAGGCGGTG TGGCTCGGAG GAACCTCTGG
2151 TAACAATGCC ACTCCCGTTC CCTGGTCAGA AAAAGCTTGC GGAGGCTAAG
2201 CACCAAGTACA GCATCGACGA GCTGAAGCAC CAGGAGAACC TGATGCAGGC
2251 CCAGCTGGAG GACGAGCAGG CGCAGCGGGA GGAGCTAGAG AAGCACAAGA
2301 CTGCGTTTGT GGAACACCTG AATGGCTCCT TCCTGTTTGA CAGCATGTAC
2351 GCTGAGGACT CAGAGGGCAA CAATCTGTCC TACCTGCCTG GTGTCGGTGA
2401 GCTCCTTGAG ACCTACAAGG ACAAGTTTGT CATCATCTGC GTGAATATT
2451 TTGAGTATGG CCTGAAACAG CAGGAGAAGC GGAAGACAGA GCTTGACACC
2501 TTCAGTGAAT GTGTCCGTGA GGCCATCCAG GAAAACAGG AGCAGGGCAA

```



```

2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAAATGC
2651 AGTGCTGACA TCAGTGAAGT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAAGTGTAGC CTTATGGGC
2801 ACGCCTCTGC TGGAAACGTC CAGCAGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCTCTG TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAGATGAT TCTTGTGGAA AATGAGAGGT GAGCTCATTC TTCTGAAATG
2951 GTCCCCCTAT CCTGGAAGTC AGTGGGGAGA GGTTTTTGAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCACT TGGGGACCAA
3051 ACAAAGACCC CTCCGCATTG CAGCCTGCAG TTGCGCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAATGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAGTCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCAAGT GATAGCACCT AATCAATTC TCACACGTC AGTGACACCT
3301 CAGACATGGG ATAGATTTC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAACA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCAGC CTTTTCAGGC ACGGAGATG CTTTCAACA
3451 TCCAAACATT TCCCAGAACC CATGTGCCAT CCTACTTGTA TTACTGGTGG
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCT ATTTTATTC
3551 ACGAAGACAG CACATACATG TGTTTGA AAAA TTATGTGAGG TGCTCACTCT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCTGCCCC ACCTTGACG
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCC
3701 TGGAGATCAC ATCTTCCATT GTTCTTCAA TCAACACTAA CTCCCATTG
3751 GGCCTTAGGT GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CCAGCCCGCT CCTACTGAGG GCCTCCTCTC TGTCAGGCAC
3851 CTTGCCAAAGC ATTTTGTGTG AAGTGACTCA TTTAACCTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 GCCCCAAGGAG GTGAAATGCC TTTCCAGAG TTACACAGAG TGCTGGAGCT
4001 GGGAAATCTG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TCTCACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACCAGG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTCACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGTCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAAGTGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAAACA TGTTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTTT CGAAAATGTC CAAAGCCTAT ATCCTTCTG TGATGACCTT
4501 CCCCATGGGG AGGTGCTACA GAGCCCTGG GCTTGTCCCG GCCTCTGGAC
4551 AAAAGAATGT TCCACAGGGT CTGAGGAGGT TTCCCGACCC TCAGAACAT
4601 GATGGCCTGG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ACTGTTTGCA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTT
4701 ACTAGGAAAT GACGCCCCCT GTGTGTCAGG CAAGCACACT CTGGGGTTGA
4751 GGCAACCCCC ACGTGGAAAG CACTATAAAG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCTT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGCCAGGAG CCTCCATGCA GGAGGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGCCC
5101 AGGGTTCGGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TCGCCTTGGG CCCTTAAAGT CTGTCTCCCG CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGGC CTGCTCCTCC CGGGCCTTTG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCCTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCCTAAC GACCTGCGCG CGCTTTTGTG CGATAAAGAT ACGATTGTTA
5401 ATGCTGTCGG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCCG AATCAACTCT TGGTGTACAC GTTTAATAGA
5501 CAGGATTCAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCGACATCC TAGACTAGAT GAATGTCAGC CACAGGAGCT TCTTCAAAC
5651 ATAGCACCAG CCCAGCCAG GAGAAGGAAG TGCACACGCC TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATCT
5751 TCCCCACCC CTGGAAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS1292248 from database EMBL:
human STS SHGC-53917.
Score = 874, P = 3.3e-33, identities = 180/185

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225
 Category: similarity to known protein

```

1 MNQPCNSMEP RVMDDDLKL AVGDQGPQEE AGQLAKQEGI LFKDVLSQL
51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGLENLAHLV WLDLSFNNIE
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
151 YLRREFKCLRT LSLSRNPIS AEDYKMFICA YLPDLMYLDY RRIDDHTASV
201 SLSVSQPCET DSSSPQVSWK RGIEE

```

BLASTP hits

Entry S68209 from database PIR:
 sds22 protein homolog - human >TREMBL:HSSDS22MR_1 gene: "sds22";
 product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA
 Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:
 suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)
 >TREMBL:SPSDS22_1 gene: "sds22+"; S.pombe sds22+ gene, complete cds.
 Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:
 protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)
 >SWISSPROT:SD22 SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT
 SDS22. >TREMBL:SPAC4A8_12 gene: "sds22"; product: "phosphatases pp1
 regulatory subunit"; S.pombe chromosome I cosmid c4A8.
 Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2_5 from database TREMBL:
 gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.
 Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

Alert BLASTP hits for DKFZphut1_20ml1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_20ml1, frame 1

Report for DKFZphut1_20ml1.1

```

[LENGTH]      225
[MW]           25955.87
[pI]           4.63
[HOMOL]        PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
[FUNCAT]       01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
YJL005w] 3e-05
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
[EC]           4.6.1.1 Adenylate cyclase 2e-06
[PIRKW]        nucleus 5e-16
[PIRKW]        duplication 2e-06
[PIRKW]        tandem repeat 2e-06
[PIRKW]        cAMP biosynthesis 2e-06
[PIRKW]        glycoprotein 2e-06
[PIRKW]        phosphorus-oxygen lyase 2e-06
[SUPFAM]       leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM]       fibromodulin 3e-07
[SUPFAM]       yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM]       yeast adenylate cyclase 2e-06
[PROSITE]      CK2_PHOSPHO_SITE 2
[PROSITE]      PKC_PHOSPHO_SITE 1

```


[KW] All_Alpha

```
SEQ  MNQPCNSMEPRVMDLKLAVGDQGPQEEAGQLAKQEGILFKDVLSLQDIFRNILRIDN
PRD  cccccccccccccchhhhhccccccccchhhhhhhhhhhchhhhhhhhhccccccccccccc

SEQ  LWQFENLRKLQLDNNIIIEKIEGLENLAHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLQVLSLGNRRIDNMMNIIYLRRFKCLRTLSSLRNPISEAEDYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccchhhhhhhhhccccccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVSLSVSQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccccc
```

Prosites for DKFZphut1_20ml1.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1_20ml1.1)

DKFZphutel_20m24

group: metabolism

DKFZphutel_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits
Alg9 is involved in the assembly of the core oligosaccharide
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)

HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```

1  TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51  TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCGGCG TGCGGACAAG
101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GCGGGCGCGG AGCACCGGAC
151 CGAGTTATCT GGGAAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGTCTGCTT TCAGCAAGGT TATGTGCTGC TCTCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTGCGTC CTATGCTTAC CTGTGCTTTC ATGCCTGGCC AGCTGCATTT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTGTGTGTTT ACTTTTTCGG
451 ATGTCTTCTG GCTTTTGTGA GCTGTATTG TGAACCTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATTGCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATGCCTTTG ATTGTCTGGT
751 CATGAACAC AGGTGGAAGA GTTCTTTTCA TTGGTCGCTG ATGGCCCTCA
801 TACTATTTCT GGTGCCTGTG GTGGTCATTG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTACTCCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCCTG GTATTTCTAT TTAATTAATG
951 GATTTTCTGA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGTGTCAG AGATTTTCATG TTCAGAATTT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTTCTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCTTTT CCCTGTGTAT
1151 CCACCTTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTGTGCAATT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTTGTA
1351 FCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGTA TCGATTTCCC
1451 AGCAGCTTCC TTCTTCTGTA CAATTGGCAG CTTCAAGTTA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGAG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTATC CAATAAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATC CTGTGCTTCT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCTCT CCTGTGATG CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAAACCC GGAAGCAAAA GCAAAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATCGCTGT AACATTTGTA
1951 ATAAAGTCT TCTGACATGA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HSAC381 from database EMBL:
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo
sapiens genomic clone cSRL-50A3.
Length = 601

Medline entries

96293493:

Stepwise assembly of the lipid-linked oligosaccharide in the
endoplasmic reticulum of *Saccharomyces cerevisiae*:
identification of the ALG9 gene encoding a putative
mannosyl transferase.

Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611
Category: strong similarity to known protein

```

1 MASRGARQRL KSGGASSGDT APAADKLREL LGSREAGGAE HRTLSGNKA
51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEGF
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVEY FLRCLLAFVS
151 CICELYFYKA VCKKFGHLHVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPFSS AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVFL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVLDL TMRETPREP
551 YSSNKEEWIS LAYRPFLDAS RSSKLLRAFY VPFLSDQYTV YVNYTILKPR
601 KAKQIRKKS G

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20m24, frame 2

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME
II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces
cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME
II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces
cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME
II.

Length = 653

HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query: 48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPHTHYLIYGEGFQTWEYSP 107
      N W + FK LLS R+ A+ I+DCDE +NYWEP H +YGEFGQTWEYSP
Sbjct: 43 NNPDNDWPFSSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFGQTWEYSP 102

Query: 108 AYAIRSYAYLLHAWPAAFHARILQTNKILVEYFLRCLLAFVSCICELYFYKAVCKKFG 167
      YAIRSY Y+ LH PA+ A+ KI+VF +R + + E Y + A+CKK +
Sbjct: 103 VYAIRSYFYIYLHYIPASLFANLFGDTKIVVFTLIRLTIGLFCLLGEYAFDAICKKINI 162

Query: 168 HVSRRMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
      R + F + S+GMF +S+AF+PSSFCM T + + + + + VA ++GW
Sbjct: 163 ATGRFFILFSIFSSGMFLASTAFVSSFCMAITFYILGAYLNENWTAGICVAFSTMVGW 222

Query: 228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```


Pedant information for DKFZphutel_20m24, frame 2

Report for DKFZphutel 20m24.2

```

[LENGTH]      611
[MW]           69863.78
[pI]           8.91
[HOMOL]        SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-
93
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69
[FUNCAT]       01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c]
4e-69
[FUNCAT]       01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69
[PIRKW]        glycosyltransferase 9e-68
[PIRKW]        transmembrane protein 9e-68
[PIRKW]        hexosyltransferase 9e-68
[PROSITE]      MYRISTYL 9
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 7
[PROSITE]      PKC_PHOSPHO_SITE 6
[PROSITE]      ASN_GLYCOSYLATION 2
[KW]           TRANSMEMBRANE 7
[KW]           LOW COMPLEXITY 6.71 %

```

```

SEQ MASRGARQRLKSGASSGDTAPADKRLRELLGSLGSGAAGAEHRTLSGNKAGQVWAPEGST
SEG .....
PRD cchhhhhhhccccccccccccchhhhhhhhhccccccccccccccccccccccccch
MEM .....MMMMMM

```

```

SEQ AFKCLLSARLCAALLSNISDCDETfNYWEPThYLIYGEgFQTWEYSPAYAIRSAYLllH
SEG ...xxxxxxxxxxxxxx
PRD hhhhhhhhhhhhhhhhhhhhhccccceeeccccceeeeeeccccceeecccchhhhhhhhhhh
MEM MMMMMMMMMMMMMMMMMMM.....M

```

```

SEQ AWPAAFHARILQTNKILVFYFLRCLLAFVSCICELyFYKAVCKKfGLHVSrMMLAFLVLs
SEG .....
PRD cchhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

```

SEQ TGMFCSSSAFLPSSFCMYTTIAMTGWYMDKTSIAVLGVAAGAILGWPfSAALGLPIAFD
SEG .....xxxxxxxxxxxxxx
PRD cceeeccccccccchhhhhhhhhhhhhccccccccccccceeeehhhhhhhccccceeeecchhhh
MEM .....MMMMMMMMMMMMMMMM

```

```

SEQ LLVMKHRWKSFFHWSLMALILFLVPVVVIDSYyYgKlVIAPlNlVLyNVfTPHGFDlyGT
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhheeeeeeeccccccccccccceeeeeeecccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

```

SEQ EPWYFYLINGfLNFNVAfALALVLPLTSLMEYllQRfHVQNLGHPyWLTlAPMyIWfII
SEG .....xxxxxxxxxxxxxx
PRD cceeeeeeccccchhhhhhhhhhhhhcchhhhhhhhhhhccccccccceeeehhhhhhhhh
MEM .....MMMMMMMMMMMMMMMMMMMM

```


Prosites for DKFZphute1_20m24.2

(No Pfam data available for DKF2phut1_20m24.2)

DKF2phutel_21d15

group: uterus derived

DKF2phutel_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGGGGGCCT GGCGGCCGTT
251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC CGCGGCCCTT CCCCTGGGCG
301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GGCGGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCCTCGA GGCCGCGAGT GGCGCGGCCA
401 GACCACTGCC AGGCTCAGCG GCGGCGCGGG CTGGGCGACG GCGAGGACGC
451 ACCGGTGGCT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTT GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCCG
601 GCCCCAACAC CGTGCCCAAG GCCCGGGGCC CGAGCCCACT TTAGGTCCCC
651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
701 CCAGGCCCTG CACGGCTGAG CCCGAGAGGA CCGGCGCTCA GCCCGGGTCC
751 CCACGCTGCC CCGGCGCGTG CTCTGCGTCG GTCCGCGCGG CTCCCACTCA
801 CTCGCTGCTG GTCGCTCTCC GGGCGGGGGC GACTTGGCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCCTCAGCT CAAGCCGCTG CTCTTCGAAA TCCCGGGCTT CCTGACTGAT
1001 GAAGAGTGTC GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTACAGCG
1051 CAGCCAGATC CTGCCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTGAGCCA GCTGGACCTC TTCCGGGTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCCGTG
1301 AGCCTCCTGC CCACCTCCAG GTGCACAATT TTGAAAACCT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTGG GTGATGGTCT CGAGGGCAGT TCTTGAGAC
1501 CCTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCCT GCTCCATCCA TCTCTCCAGC
1651 CAGACACAGG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCTT TCCTCTTGTC CACCTGGAGT CATGCCGAAG CGCCTAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGA AGGTGATGTC CTGGGTGTCG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTGTTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTTGC TAAGTGGGCT GGTAAACGAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCCC AGGATGAAAC CTGGGGTCAT GAGGAACCTC
2301 CCGGGGGCTG GCCCTGCTTG CACCCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCCAAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCCGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGTCCT TGGCCATATG
2451 GCTCAAGCTT CCTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAACCTGAG AAAAGGAACT
2551 GCTCTGGGCT TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTA CAAGGCCCTC CAGTGCCTGA
2651 GGTCAAGGGC TGCTGCTGCC TTTGGGGTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCT CCAATATGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTCTC CAACATGGAC CTTCCGGACT
2801 TCCACAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGCGG
```



```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGCCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGTCCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAATGA CCTCTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCCGCC CCCTTTATTT AGTGGAAATG TCAACATTTT
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCTTGTCT CCCACCCAGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCGG CTGCAGGTG TTGATATGG TGAGGGGGGC
3601 CACTACCATT CCCACGTGGA CAGTGGGCC GTGTACCCAG AGACCATCTG
3651 CTCCCATACC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCCACAAGTT GTTTACCCAA TGGCTGGAAG GGGGTGGCTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAACTTG
3951 CCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCTTA GCTTGCTCTG
4001 CCCATTCTCT CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCCCTCCCTC CCCTGTCAAG CCAGTTCTTT CCTCTTCAGG TGGCTGTCTT
4101 GGGCCAGCCC CTTCCCATCC CCAAGGAGCC CTTAGCGCGC CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCCC TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGGCCTATT GTCTCCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CCTGGGCAGC TTATCCTGCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCT CCCCTGCCTT ACAGCTACAT GACAGTGTCTG
4451 TTTTATTTGA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTGTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAAT ATTCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTCAACC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCACCCT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGGAAA CCTGCGTGTG AAGCCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACCTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGGATTGCCA ACAACTGGAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCAA CAGGAGATGG CCCGCTTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCCA GCCCGGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAGGTTCT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCCCTAT ATTCAATGTTA TTTATTGTGT ACTGACTCCA TCTGCCCGCT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAA GG

```

BLAST Results

Entry HSU64252 from database EMBL:
 Human STS sequence NOTI-225.
 Score = 959, P = 1.2e-36, identities = 195/199

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118
 Category: questionable ORF
 Classification: no clue

```

1 LPLVYALMVP LLSASTLGLT ASDLESVQLC PTATQLGKRS PSVGWGSRRR
51 KAEFGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRLPWA
101 RARPGCHGGS GDRPAA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 1

No Alert BLASTP hits found

Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191
Category: putative protein
Classification: no clue

1 MAAAVTGQR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPAAAPGA
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAAARCAP P

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1
Length = 1,298

HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 36/103 (34%), Positives = 44/103 (42%)

Query: 87 GDESSDPGPQHRAQGPPEPTLGLPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
G + PGP G GP P P T+ G S R P PA S P GP +P
Sbjct: 726 GRKRKSPGPAPRPPGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARS GAPAAARCAP 189
AAP AA ++R P+ GP LG W + P+ AP
Sbjct: 783 PAAPRAAAQARPRPVAVSRRPAEGPDPLGG-WRRQPPGPSHTAAP 827

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 8/21 (38%), Positives = 9/21 (42%)

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48
DH + A G G AP P
Sbjct: 212 DHAREARAVGRGPSSAAPAP 232

Pedant information for DKFZphut1_21d15, frame 1

Report for DKFZphut1_21d15.1

[LENGTH] 117
[MW] 11797.32
[pI] 10.68
[KW] Irregular
[KW] SIGNAL PEPTIDE 22
[KW] LOW_COMPLEXITY 38.46 %

SEQ LPLVYALMVPLLSASTLGTASDLESVQLCPTATQLGKRSPSVGWGSRRRKAEPGADAGG
SEGXX
PRD cccccccccccccccccchhhhhhhcccccccccccccccccccccccccccccccccc

SEQ SGRAQHQPAPSPSPDRGARGPGGRCPGDCAARAPPRPLPWARARPGCHGGSGGDRPAA
SEGXX
PRD ccc

(No Prosite data available for DKFZphut1_21d15.1)

(No Pfam data available for DKFZphut1_21d15.1)

Pedant information for DKFZphut1_21d15, frame 2

Report for DKFZphut1_21d15.2

{LENGTH} 191
{MW} 19916.88
{pI} 10.43
{KW} TRANSMEMBRANE 1
{KW} LOW_COMPLEXITY 29.84 %

SEQ MAAAAVTGQRPETAAAEASRPQWAPPDHCQAAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG
PRD cccceeeccccchhhhhhhhhccccccchhhhhhhccccccccccccccccccccchhh
MEM

SEQ FLVLMVFVHLYLGNVLALLLFVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEGxxxxxxxxxxxxxxxx.....
PRD hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeee
MEMMMMMMMMMMMMMMMMM.....

SEQ TSLPRRAPGPAPRLSPRGPAALRRSRALPLTRLRLSLSGPGRGPFWAARS
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....xxxx
PRD eccccccccccccccccccccccccccccchhhhhhhccccceccccccccchhhhhc
MEM

SEQ GAPAAARCAPFP
SEG xxxxxxxxxx..
PRD ccccccccccc
MEM

(No Prosite data available for DKFZphut1_21d15.2)

(No Pfam data available for DKFZphut1_21d15.2)

DKFZphutel_22d2

group: signal transduction

DKFZphutel_22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp

Poly A stretch at pos. 3230, no polyadenylation signal found

```
1 CTCCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
151 CCAGAAGAGG TTCCTCCCCG GGCAGAAGAA ATCACCATTG CAGCTGATGT
201 CACCCACAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCATCTGT
301 ATAGTGATAT CCGTTAACAA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGAATATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTACAGC AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGTACTGCCC AGAGGAGAAG
601 GAGATGA AAC CAGCTTGTAT AAAAGCCCTT ACTCGTATAT TTAATAATATC
651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAACCT AACTTCTTTC
701 AGAGGATTGG TTTCACACT CCATTAGCTC CTCAAGCTCT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCTGAAA GGTTCCTCT TTTACACAC ACTTTTATC CAGAGAGGGA
851 GACACGAAAC TACTTGGACT GTGCTTCGAC GATTTGGTTA TGATGATGAC
901 CTGGATTGTA CACCTGAATA TTTGTTCCCC CTGCTGAAAA TACCTCCTGA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTGGAT AGAGACTGTG CTTTGTACCC TGATGAGCTT
1051 AAAGATTTAT TTAAAGTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCAGTGG GACGCTCAG ACTTATTAG ATGTACAGCG GTGCTTGGAA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTCAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAACCTCAAG AATGTGTTT AGATGTAATG TAATTGGAGT GAAAAACTGT
1351 GGGAAAAGTG GAGTCTTCTA GGCTCTTCTT GGAAGAAACT TAATGAGGCA
1401 GAAGAAAAAT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTGTT TGCTATGATAT CTCAGAAATCG
1501 GAATTCTTAA CTGAAGCTGA AATCATTGTT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAAAT CCTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTCACCTA CTGATTTCTG
1701 CAGGAAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGCTGGTAT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAATT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGGAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCAC CCAGCATTAA AAAATAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTT CTGTTTTTGA CCTAGGTAT ATGAAGTTT
2201 CTAAATATTT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCAAGTTAG
2251 GATATGTCTT TTTTAAAGTG TGTAAGAGAT AGTTGTAATT GGAATTTCTA
2301 CTGTATATAA GTTTTACATT AAGTGTTACG AGCCACAAT TFCATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAATTTG CATGGTTCTT AATGGCATTG
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTG TTTTGTAAAC AAATAGTTT
2501 TCAGGCAGTG CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG
```



```

2601 AGTAATAATC AACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTTGATTG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTTGTAT
2751 TCGAAGACAA GTCATTCTTA TTATTATAGA ATAACCAAAA CCTTATTTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTCTGTGGG ATTTGTGTA TATTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCTTAC
3101 TACATAAATG TACTTCTTTA ATCTGTGAT TACAGTAACT GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAAATCTG TTGTGTGTGT
3201 GTGTGTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAA

```

BLAST Results

Entry AC004527 from database EMBL:
 *** SEQUENCING IN PROGRESS *** NF1-related locus, Direct Submission;
 HTGS phase 1, 10 unordered pieces.
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HS148355 from database EMBL:
 human STS SHGC-31220.
 Score = 1826, P = 7.5e-78, identities = 388/406

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 64 bp to 1803 bp; peptide length: 580
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSI IMSLVSEEF EEVPPRAEEI TIPADVTPER
51 VPTHIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPIMNQYTE IETCVECSAK
151 NLKNISELYF YAKKAVLHPT GPLYCPEEKE MKPACIKALT RIFKISDQDN
201 DGTLNDALN FQRICFNTP LAPQALDVK NVVRKHISDG VADSGTLKLG
251 FLFLHTLFIQ RGRHETWTV LRRFGYDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHDLDL DCALSPDELK DLFKVFYPIY WGPDVNNTVC
351 TNERGWITYQ GFSLQWTLTT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDKL KKQTQNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVYVYGQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSEYECARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPQAFATCN TADAPSKDIF VKLTTMAMYP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22d2, frame 1

TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320_4 gene: "SPCC320.04c"; product: "hypothetical protein";
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11.
 Length = 625

HSPs:

Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138
Identities = 263/582 (45%), Positives = 380/582 (65%)

Query: 4 DVRILLVGEPRVGKTSLIMSLVSEEFPEEVPRAEITIPADVTPERVPTHIVDYSEAEQ 63
DVRIL+L+G+ GKTSL+MSL+ +E+ + VP R + + IPADVTPE V T IVD S E+
Sbjct: 9 DVRIVLIGDEGCGKTSVMSLLEDEWVDAVPRRLDRVLIPADVTPEVNTTSTVDLSIKEE 68

Query: 64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
+ + EI QANVIC+VY+V ++ ++D + ++W+PLI + + P+ILVGNKSD
Sbjct: 69 DENWIVSEIRQANVICVYVSVTDESTVDGIQTKWLPILIRQSFGEYHETPVILVGNKSDGT 128

Query: 124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKMP 183
++ + ILPIM TE+ETCVECSA+ +KN+SE+FYAQKAV++PT PLY + K++
Sbjct: 129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYAQKAVIYPTRPDYDADTKQLTD 187

Query: 184 ACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALEDVKNVVRKHISDGVAD 243
KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+
Sbjct: 188 RARKALIRVFKICDRDNDGYLSDTELNDQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query: 244 SGLTLKGFLFLHTLFIQGRHETTWTVLRFRFGYDDDLDTPEYLFPLLKIPPDCTTELNH 303
L L GFL+LH LFI+RGRHETT VLR+FGY+ L L+ +YL+P + IP C+TEL+
Sbjct: 248 DSLMLAGFLYLHLLFIERGRHETTAVLRKFGYETSLKLSYDLYPRITIPVGCSTELSP 307

Query: 304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVDVNTVCTNERGWITYQGFL 363
F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++
Sbjct: 308 EGVQFVSALFEKYDEDKDGLSPSELQNLFSVCPVPVITKDNILALETNQRGWLTNGYM 367

Query: 364 SQWTLTTYLDVQRCLEYLGYLSILTEQESQAS----AVTVTRDKKIDLQKKQTQRNVF 419
+ W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF
Sbjct: 368 AYWNMTTLINLTQTFEQLAYLGFVGRSGPGRAGNTLDSIRVTRERKKDLENHGTDRKVF 427

Query: 420 RCNVIGVKNCKGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYQGEKYLHLLDI--- 476
+C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLLL ++
Sbjct: 428 QCLVVGAKDAGKTVMQSLAGRMADVAQIGRRH-SPFVINRVRVKEESKYLLLREVDVL 486

Query: 477 SESEFLTEAEIICDVLVDVSNPKSFEYCARIKQHFMSRIPCLIVAAKSDLHEVKQ 536
S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q
Sbjct: 487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQKYFYRTKTPCVMIATKVEREEVDQ 546

Query: 537 EYSISPTDFCRKHKMPPPOAFTCNTADAPSKDIFVKLTMMAMP 580
+ + P +FCR+ ++P P F+ S IF +L MA+YP
Sbjct: 547 RWEVPPEEFRCQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

Pedant information for DKFZphutel_22d2, frame 1

Report for DKFZphutel_22d2.1

[LENGTH] 580
[MW] 66541.61
[pI] 5.56
[HOMOL] TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07


```

[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
1e-07
[FUNCAT]      08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT]      10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT]      09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 9e-04
[BLOCKS]      BL00410A Dynamin family proteins
[SCOP]         dlplk_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 2e-42
[SCOP]         dlguaa_ 3.25.1.3.10 Rap1A [Human (Homo sapiens) 5e-59
[PIRKW]        transmembrane protein 1e-79
[PIRKW]        membrane trafficking 2e-06
[PIRKW]        acetylated amino end 3e-09
[PIRKW]        prenylated cysteine 3e-09
[PIRKW]        signal transduction 1e-07
[PIRKW]        transforming protein 3e-09
[PIRKW]        immediate-early protein 8e-06
[PIRKW]        alternative splicing 4e-08
[PIRKW]        P-loop 1e-10
[PIRKW]        lipoprotein 7e-10
[PIRKW]        proto-oncogene 3e-09
[PIRKW]        methylated carboxyl end 3e-09
[PIRKW]        membrane protein 3e-09
[PIRKW]        GTP binding 1e-10
[PIRKW]        thiolester bond 7e-10
[SUPFAM]       ras transforming protein 1e-10
[PROSITE]      ATP_GTP_A 2
[PROSITE]      MYRISTYL 3
[PROSITE]      EF_HAND 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 14
[PROSITE]      TYR_PHOSPHO_SITE 4
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         Ras family (contains ATP/GTP binding P-loop)
[KW]           Irregular
[KW]           3D

```

```

SEQ      MKKQDVRIILLVGEPRVGKTSLIMSLVSEEFPEVPPRAEETIPADVTPERVPTHIVDYSE
1jai-    ...EEEEEEETTTTCHHHHHHHHHHCCCCCCCCCCCCEEEEEEETTEEEEEEEECCECC

SEQ      AEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKS
1jai-    CGGGHHHHHHHHHTTEEEEEETTTTHHHHHHHH-HHHHHHHHHHCTTT-TCEEEEEETT

SEQ      DLVEYSSMETILPIMNQYTEIETCVCESAKNLKNISELFYYAQAVLHPTGPLYCPPEKE
1jai-    TTTTTTTTHHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....

SEQ      MKPACIKALTRI FKISDQDNDGTLNDAELNFFQRICFNTPLAPQALDVKNVVRKHISDG
1jai-    .....

SEQ      VADSGTLKGLFLHLTLFIQGRHETTWTVLRFRFYDDDLDTPEYLFPLLKIPPDCTTE
1jai-    .....

SEQ      LNHAYLFLQSTFDKHLDRDCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQ
1jai-    .....

SEQ      GFLSQWTLTTYLDVQRCLEYLGYLSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
1jai-    .....

SEQ      CNVIGVKNCCKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLDISESE
1jai-    .....

SEQ      FLTEAEIICDVVCLVDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-    .....

SEQ      SPTDFCRKHKMPPQAFCTCNTADAPSKDIFVKLTMMAMP
1jai-    .....

```

Prosites for DKFZphut1_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PD0C00001
PS00001	154->158	ASN_GLYCOSYLATION	PD0C00001
PS00001	346->350	ASN_GLYCOSYLATION	PD0C00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	94->97	PKC_PHOSPHO_SITE	PD0C00005
PS00005	105->108	PKC_PHOSPHO_SITE	PD0C00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

Pfam for DKF2phutel_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK		
	++L+G+ VGK++L ++ EF+EE +P ++ T ++ +++		
Query	6	RILLVGEPRVGKTSLIMSLVSEEFPEE-VPPR-AEEITIPADVTPEVP	52
HMM	LQIWDTAGQERYrSMRPMYRGAMGFMLVYDITNRqSFENIr.NWweEIr		
	I D E+ + + +A+++ +VY+++N+ S ++++ +W+++ I+		
Query	53	THIVDYSEAEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVstEEGQeFAREWGAIPFMETSAKT		
	+ D+D+ P +LVGNK+DL + ++T + +E+SAK+		
Query	103	ERTDKDSRLPLILVGNKSDLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
	N+ E F+ + +++L + +++ +++++ + C+		
Query	152	LKNISELFYIAQKAVLHPT-----GLYCPEEKEMK-PACI--	186

DKFZphut1_22e12

group: signal transduction

DKFZphut1_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits
cornichon is required for signal transduction in the EGF-receptor
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp
Poly A stretch at pos. 499, no polyadenylation signal found

```

1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTCGCGCG
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCAGTGGTT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTTGGAA
451 CAAAAAATA TTTTGCTGT ATTTTACCA TATAAAGTAT TAAAAAACA
501 TGAAAAAATA AAAAAAATA

```

BLAST Results

No BLAST result

Medline entries

95300228:
cornichon and the EGF receptor signaling process are necessary for both
anterior-posterior
and dorsal-ventral pattern formation in Drosophila.

Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92
Category: strong similarity to known protein

```

1 MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKWWIP
51 ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398_1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
Length = 138

HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV 60
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++
Sbjct: 1 MGAWLFILAVVNCINLFGQVHFTILYADLEADYINPIELCSKVNKLITPEAALHGALS 60

Query: 61 LLLMSLHWFIPLLNPVATWNIYRM 85
L L++ +WF+FLNLNPV +N+ ++
Sbjct: 61 LFLNGYWFVFLNLNPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90
+YRMI+ALI
Sbjct: 123 LYRMIMALI 131

Pedant information for DKFZphut1_22e12, frame 1

Report for DKFZphut1_22e12.1

[LENGTH] 92
[MW] 10614.98
[pI] 5.04
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
5e-14
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGL054c]
2e-15
[PIRKW] transmembrane protein 2e-11
[PROSITE] CK2_PHOSPHO_SITE 3
[KW] SIGNAL PEPTIDE 33
[KW] TRANSMEMBRANE 2

SEQ MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV
PRD ccchhhhhhhhhhhhhhhhhheeeccccccccccccccccceehhhhhhhhhhhhh
MEMMMMMMMMMMM

SEQ LLLMSLHWFIPLLNPVATWNIYRMILALIND
PRD hhhhhhhheeeccccchhhhhhhhhhhhhccc
MEM MMMMMMMMMMMMMMMMM..MMMMMM....

Prosite for DKFZphut1_22e12.1

PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1_22e12.1)

DKF2phute1_22n2

group: uterus derived

DKF2phute1_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMF2

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```

1 ACAACAGGCT GGTGCTTGG CGTGGAATCC TAAAGTGGCC TGGCTTTGAG
51 ACTGGAGTGA GACCCAGGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTTGAAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAAC GACTCCGAAG AGGTGCAAG TATTTATACT
201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAAACTG ATTCTGATTC
401 TGATGATCAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTTG CCAGTTTCTG CTGAAATTAA GGAACCTCTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAAACT
551 GAAGCCTTTC ATTCCTGATT TTATCCCAGC TGTCGGGGAT ATTGATGCAT
601 TCTTAAAGGT CCCACGTCCT GATGGAAGGC CTGACAACCT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCTACGG TGCTCTCACT
701 CTGGTTAACA GAGAATTCTA AGCAGCACAA CATCACAAA CATATGAAAG
751 TAAAAAGCCT AGAAGATGCA GAAAAGAATC CCAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CCGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAAGTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAAGCTG GAGACATGGA
1151 GACATTAAAC TTCAGCTGAG ACATTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCCT CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCTTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCCTGCT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CTTTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTCAATTCG AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA

```

BLAST Results

Entry HS188252 from database EMBL:

human STS WI-12265.

Score = 2554, P = 4.1e-109, identities = 556/587

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304
 Category: putative protein


```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDS
51  DDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GDIDAFKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHMVKVSL EDAENPKAID TWIESISELH RSKPPATVHY
201 TRMPDIDTL MQEWSPEFEE LLGKVSLETA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFALAE GKKAFTPSSN STSQAGDMET
301 LTFS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*)
Length = 562

HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:      3  DNSSDECEEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSSETDSDDDDDEEHGAPLEG 62
             +  DE EEE++ E++ T          +++DDDDDDDD + D D DDD++E A  G
Sbjct:     497 EEDDDDEDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

```

```

Query:      63 AYD 65
             D
Sbjct:     557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:      4  NSSDECEEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSSETDSDDDDDEE 55
             N+ +E ++E+ +E      + T + + N+DDDDDDDD + D D DDD++
Sbjct:     494 NNEEEDDDDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphut1_22n2, frame 3

Report for DKFZphut1_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION 1
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 10
[PROSITE]      PKC_PHOSPHO_SITE 1
[PROSITE]      ASN_GLYCOSYLATION 3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 11.84 %

```

```

SEQ  MADNSSDECEEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSSETDSDDDDDEEHGAPL
SEG  ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccchhhhhchhhhhcccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGDIDAFKVP
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccceec

```

```

SEQ  RPDGKPDNLGLLVLDEPSTKQSDPTVLSLWLTENSKQHNIQHMVKVSLDAENPKAID
SEG  .....
PRD  cccccccccceeeccccccccccccchhhhhccccccccccccchhhhhhhccccch

```

```

SEQ  TWIESISELHRSKPPATVHYTRMPDIDTLMQEWSPEFEE LLGKVSLETA EIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhccccceeeccccccccchhhhhccccchhhhhccccccccchhhhhhh

```

```

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFALAE GKKAFTPSSNSTSQAGDMET
SEG  .....

```


Prosites for DKFZphute1_22n2.3

(No Pfam data available for DKFZphute1_22n2.3)

DKFZphutel_22o2

group: uterus derived

DKFZphutel_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="11p15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCGCGG GTCACAAGGA GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCCTCAGGC
151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCCGG CCCCCTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGCCAGGG CGGGGCCGGC GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCCGTCCCGC CTTCCCGGCG GCGCCATGGA GCCCCGGGCG GTTGCAGAAG
351 CCGTGGAGAC GGGTGAGGAG GATGTGATTA TGGAAGCTCT GCGGTCATAC
401 AACCAGGAGC ACTCCAGAG CTTCACGTTT GATGATGCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCTCCA CCGTGTCTC TGCTGCAGA GTGTCCGAAT CCTGTCCCGG
551 GACCGCAACT GCCTGGACCC GTTCACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCCTGCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCCG GCCTAGTGGT
751 GAAGCTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCCAGC
801 ATGTCCAGTT CTTTGAAGTG CGGCTCCTCT TCCTGCTAAC GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCCTGTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGAACCCCCC
951 CACCCACGCT CCTTCCTTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGACGCT GCCCTTTACC GACACCTGGG GACCCTTCTC CGGCACTGTG
1101 TGATGATCGC TACTGCTGGA GACCGCACAG AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGGAACCT GCCCTCAAG TGCTGGATG TTCTCCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGGAGTG AATATGGATG
1251 TGATTGCTGC CCTCTCATC TTCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTGTG AGCGTGTGTA CTGAATGTGC
1351 CCGGATGCAC CGCCAGCCA GGAAGTTTCT GAAGGCCAGG GGATGGCCAC
1401 CTCCCCAGGT GCTGCCCTCT CTGCGGGATG TGAGGACACG GCCTGAGGTT
1451 GGGGAGATGC TCGGAACAA GCTTGTCCGC CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTTCTT GTTGTCTCTG TGCTCTGAGA
1551 GTGTGCCCGG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCCTCAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCTGTG GACCGGGAGG GTGGAGGAGA AGCCGCTTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCACGAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCGGGGGTCA TCTTACGTCC CTGCAGGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTCTGCTCC CCCATCAGGA CTGGTGCTGC TTCCAGAGAC TTCTTGGGG
2001 TTGCAACCTG GGAAGCCAC ATCCCACTGG ATCCACACC GCCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CCGTTCTGTT CATGATTGTC
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTTCA AAGCTGCTGC CTGAGGGCAG GGCCACCTG
2251 GGCACACAGA AGAGCATATG GAGGGGACAG GGTGTTGGTG TGGGTGCACA
2301 CAAAGCAAGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACTTCCCA GGGAGAAAAC CTGTGAGAAC TTTCCATACG
2401 AGTATATCAG AACACACCCT TCCAAGGTAT GTATGCTCTG TTGTCTCTGT
2451 CCGTCTTCCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCTTGG
2501 TCCTCGTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAACCTTCC TTCTACTGAG TTTAGACTAC ACTTACAACA
2601 AGGAAAATGC CCTCTGTGTG ACCATAGATT GAGATTTATA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```


BLAST Results

Entry AF015416 from database EMBL:
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:
human STS SHGC-15914.
Score = 1143, P = 9.0e-46, identities = 245/255

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEH S QSTFDDAQQ EDRKRLAELL
51 VSVLEQGLPP SHRVIWLQSV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSPVESAD MDVVLES LKC LCNLVLS SPV AQMLAAEARL VVKLTERVGL
151 YRERSFP HDV QFFDLRL LFL LTALRTDVRQ QLFQELK GVR LLTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVL FN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCV M IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPLRDVR TRPEVGEMLR NKLVRMLTHL DTDV KRVA AE
401 FLFVLCSESV PRFIKYTG YG NAAGLLAARG LMAGGRPEGQ YSEDEDTDID
451 EYKEAKASIN PVTGRVEEKP PNPMEGMT EE QKEHEAMKLV TMFDKLSRNR
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDDP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phutel_22o2, frame 2

TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7.
Length = 362

HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query: 215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRLRHCVMIATAGDRTEEFHG 273
      SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct: 12 SQDNEMVLTEILRLLEFPISKRSYLKEEDEQKILL-----LVIEIWASSLNNPNNSPLRW 65

Query: 274 HAVN-LLG-NLPLKCLDVLTLLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH---RL 327
      HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct: 66 HATNALLSFNLQLLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query: 328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPLRDVRTRP-EVGEMLRNKLVR 386
      + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct: 122 QNTLPPIAAILLSLLSFFNIKQNL-----SMLLFPTNDDRQSLQKGKSFRCLLRL 173

Query: 387 MT-HLDTDVKRVA AEFLVLCSESVPRFIKYTG YGNAAGLLAARGLMAGGRPEGQYS--- 442
      +T + + A L LC + + G G A G+ M P + +
Sbjct: 174 LTIPIVEPIGTYYASLLNELCDGDSQQIARIFGAGYAMGISQHSETMPFSPLSKAASPV 233

Query: 443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMT EEQKEHEAMKLVMTFVKLSRN 499
      + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct: 234 FQKNSRGQENTENNLAIDPITGSMCTNRNKSQRLE-LSQEEKEREAEERLFYLFQRLEKN 292

```


Query: 500 RVIQ 503
 IQ
 Sbjct: 293 STIQ 296

Pedant information for DKFZphutel_22o2, frame 2

Report for DKFZphutel_22o2.2

[LENGTH] 537
 [MW] 60372.53
 [pI] 5.20
 [BLOCKS] BL00415L Synapsins proteins
 [PROSITE] MYRISTYL 4
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 9.50 %

SEQ MEPRVAEAVETGEEDVIMEALRSYNQHSQSFTFDDAQQEDRKRLAELLVSVLEQGLPP
 SEG
 PRD cccchhhhhhhhhccchhhhhhhhhccccccccceccchhhhhhhhhhhhhhhhhhhccccc

SEQ SHRVIWLQSVRILSRDRNCLDPFTSRQSLQALACYADISVSEGSVPESADMVSVLESKLC
 SEG
 PRD cceeeccccccccccccccccchhhhhhhhhhhceeeccccccccchhhhhhhhh

SEQ LCNLVLSPPVAQMLAAEARLVVKLTERVGLYRERSFPHDVQFFDLRLFLLTALRTDVRQ
 SEG
 PRD hhhhhccccchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhh

SEQ QLFQELKGVRLLTDTLELTGVTPEGNPPPTLLPSQETERAMEILKVLFNITLDSIKGEV
 SEG
 PRD hhhhhhhchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhccccchhh

SEQ DEEDAALYRHLGTLRLHCVMIATAGDRTEEFHGHAVNLLGNLPLKCLDVLTLLEPHGDST
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccceeecccccccc

SEQ EFMGVNMDVIRALLIFLEKRLHKTHRLKESVAPVLSVLTECARMHRPARKFLKAQGWPPP
 SEG
 PRD eeehhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeehhhhhhhhhhchhhhhhhhhcccccccc

SEQ QVLPLRDVTRPEVGEMLRNKLVRMLTHLTDVQRVAAEFVLCSSESVPRIKYTGYG
 SEG
 PRD cccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccccccceeecccc

SEQ NAAGLLAARGLMAGGRPEGQYSEDEDTDTDEYKEAKASINPVTGRVEEKPPNPMEGMTEE
 SEG
 PRD chhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccccccccceeeccccccccchhh

SEQ QKEHEAMKLVTMFDKLSRNRVIQPMGMSPRGHLTSLQDAMCETMEQQLSSDPDSDPD
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhcccccccc

Prosite for DKFZphutel_22o2.2

PS00001	230->234	ASN_GLYCOSYLATION	PDOC00001
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	69->72	PKC_PHOSPHO_SITE	PDOC00005
PS00005	84->87	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	145->148	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	324->327	PKC_PHOSPHO_SITE	PDOC00005
PS00005	463->466	PKC_PHOSPHO_SITE	PDOC00005
PS00005	508->511	PKC_PHOSPHO_SITE	PDOC00005
PS00006	12->16	CK2_PHOSPHO_SITE	PDOC00006
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_22o2.2)

DKFZphutel_23e13

group: metabolism

DKFZphutes3_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1  GGTATTATTA  GCTCCTGGCT  CCGCTCTAGA  CCTCAGCGGT  TCTGGCTGCC
51 AGCCTGGGCA  CCCTGGGAAG  CCTGGGAGGA  CGGTGGCTTG  CCGGTCTGTC
101 GTGAGGCAGT  GCGGACGGGG  ACCCTCTGGG  ATTCTGCTGG  ATCTGCCCCG
151 GGGGTTACCT  TTGGGGGCTG  GGACCCAGT   CGAGGGGACA  CAACCGTCCC
201 TGGCAGTGGT  TGGTTCTGCT  TCTCCCTGCA  GAAAAGCAGC  ATTTTCGGAA
251 GCTGAAGAA  AAGCTAGCCC  AGCCACACCA  CCTTGTGTG   TGACCTTGGG
301 CAGGTGGTTC  TGTCTCTCTG  AGCCTCTGTT  TCTCTCTGAG  CTGAGCAGCC
351 ACCATGGCTG  ACGGTTCAGT  GCCCTTCTCC  TGCCACTACC  CAAGCCGCCT
401 GCGCCGAGAC  CCCTTCCGGG  ACTCTCCCTC  CTCCTCTCGC  CTGCTGGATG
451 ATGGCTTTGG  CATGGACCCC  TCCCAGACG   ACTTGACAGC  CTCTTGGCCC
501 GACTGGGCTC  TGCCCTCGTC  CTCCTCCGCC  TGGCCAGGCA  CCCTAAGGTC
551 GGGCATGGTG  CCCCAGGGCC  CCACATGCCA  CGCCAGGTTT  GGGGTGCCCT
601 CCGAGGGCAG  GACCCCCCA  CCCTTCCCTG  GGGAGCCCTG  GAAAGTGTGT
651 GTGAATGTGC  ACAGCTTCAA  GCCAGAGGAG  TTGATGGTGA  AGACCAAAGA
701 TGGATACGTG  GAGGTGTCTG  GCAAACATGA  AGAGAAACAG  CAAGAAGGTG
751 GCATTGTTTC  TAAGAACTTC  ACAAGAAAA   TCCAGCTTCC  TGCAGAGGTG
801 GATCCTGTGA  CAGTATTTGC  CTCACCTTCC  CCAGAGGGTC  TGCTGATCAT
851 CGAAGCTCCC  CAGGTCCCTC  CTTACTCAAC  ATTTGGAGAG  AGCAGTTTCA
901 ACAACGAGCT  TCCCCAGGAC  AGCCAGGAAG  TCACCTGTAC  CTGAGATGCC
951 AGTACTGGCC  CATCCTTGT  TTGTCCCCAA  CCCTAGGGCT  TCTCTGATTC
1001 CAGGATACAT  TACTTTAGCT  GAACTCAGAT  TTAGTGCAAG  TAAAATGTTA
1051 GAGGTGCGG  GGGTGAGGAC  TGACCACAGA  TTCCCTGGAT  AGTGTAGTGG
1101 TAGATTTCTC  CACAGGATAG  CGCAATTGGC  AAATCATGCT  TGGTTGTGTT
1151 AGGCCAAAAT  ACTAGTTTGT  CTTTCTTTAC  CTTTCTATC   TTGATGAAAA
1201 TGTGTGCACAT  TCTATAGTTG  CAAACACAT   AAAAGGGGAC  TTAACATTTC
1251 ACGTTGTATC  TTAATTGCAG  TGAATGCAAG  GGTTACTTTT  CTCTGGGGAC
1301 CTCCCCATC  ACCCAGGTT  CTAATCTGGG  CTCCCAGTTC  CCATGGCTCC
1351 CAAACCATGC  CGCATGGTTT  GGTAAATGAA  ACCCAGTAGC  TAACCCCACT
1401 GTGCTTCCAC  ATGCCTGGCC  TAAATGGGT   GATATACAGG  TCTTATATCC
1451 CCATATGGAA  TTTATCCATC  AACCAATAA   AAACAAACAG  TGCCTTCTGC
1501 CCTCTGCCCA  GATGTGTCCA  GCACGTCTC   AAAGTTTCCA  CATTAGCACT
1551 CCCTAAGGAC  GCTGGGAGCC  TGTCACTTTA  TGATCTGACC  TAGGTCCCCC
1601 CTTTCTTCTG  TCCCCTGTGT  TTAAGTCGGG  ATTTTACAG   AGGGAGCTGT
1651 CTCCAGACAG  CTCCATCAGG  AAACAAGCAA  AGGCCAGATA  GCCTGACAGA
1701 TAGGCTAGTG  GTATTGTGTA  TATGGGCGGG  ACGTGTGTGT  CATTATTATT
1751 TGAGTTATGC  TGTGTTTAG  GGGTAAATAA  CAGTAAATAA  TTAATAATAA
1801 TAATAATAAT  AATAAAGGAG  CTGACGTTCT  TAAAAAAGAA  AAAAAAATAA
1851 AAAA
```

BLAST Results

Entry HS286348 from database EMBL:
human STS TIGR-A002J47.
Score = 510, P = 1.2e-16, identities = 102/102

Medline entries

95394379:
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:
Physiological and pathological changes in levels of the two small stress proteins, HSP27 and alpha B crystallin, in rat hindlimb muscles

Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196
Category: strong similarity to known protein
Prosites motifs: SUBTILASE_ASP (28-39)

```

1 MADGQMPFSC HYPSRLRRDP FRDSPLSRL LDDGFGMDPF PDDLTA SWPD
51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGEPWKVCV
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P = 4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561_1 product: "heat shock protein HSP27"; Mus musculus heat shock protein HSP27 internal deletion variant b mRNA, complete cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog
Length = 209

HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27
Identities = 80/182 (43%), Positives = 102/182 (56%)

```

Query:      1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRL LDDGFGMDPF PDDLTA SWPDWALPRLSS 58
            M + ++PFS  PS  DPFRD  P  SRL D  FG+  P++  W  W  S
Sbjct:      1 MTERRVVFSLRLSPSW---DPFRDWYPASRLFDQAFGLPRLPEE----WAQWFG---HS 50

Query:      59 AWPGTLRSGMVP---RGPTATARFGVPAEGR--TPPPFPG-----EPWKVCVNVHSF 105
            WPG +R  +P  GP A A  PA R  +  G  + W+V ++V+ F
Sbjct:      51 GWPGYVRP--IPPAVEGPAAPAAAAAPAYSRLSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query:      106 KPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLI 165
            PEEL VKTKDG VE++GKHEE+Q E G +S+ T K LP VDP V +SLSPEG L
Sbjct:      109 APEELTVKTKDGVVEITGKHEERQDEHGYISRLTPKYTLPPGVDPTLVSSSLSPGTLT 168

Query:      166 IEAPQVPPYSTFGE 179
            +EAP  P +  E
Sbjct:      169 VEAPMPKPATQSAE 182

```

Pedant information for DKFZphut1_23e13, frame 3

Report for DKFZphut1_23e13.3

[LENGTH] 196
[MW] 21604.37


```

[PI] 5.00
[HOMOL] PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS] BL01031C
[PIRKW] blocked amino end 1e-13
[PIRKW] acetylated amino end 4e-13
[PIRKW] phosphoprotein 7e-21
[PIRKW] glycoprotein 2e-11
[PIRKW] heat shock 7e-21
[PIRKW] molecular chaperone 4e-13
[PIRKW] alternative splicing 1e-19
[PIRKW] eye lens 6e-14
[PIRKW] stress-induced protein 7e-21
[SUPFAM] alpha-crystallin 7e-21
[PROSITE] SUBTILASE_ASP 1
[PROSITE] MYRISTYL 2
[PROSITE] CK2_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 6
[PROSITE] ASN_GLYCOSYLATION 1
[PFAM] Heat shock hsp20 proteins
[KW] All_Beta
[KW] LOW_COMPLEXITY 7.14 %

```

```

SEQ      MADGQMPFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTA SWPDWALPRLSSAW
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccccccccccccccchhhhhcccccccccccccccccccccccccc

SEQ      PGTLRSGMVPRGPTATARFGVPAEGRTPPPFPGEPWKVCNVVHSFKPEELMVKTKDGYVE
SEG      .....
PRD      cccccccccccccchhhhhhhccccccchhhhhhheeeeeccccccceeeeeecccccee

SEQ      VSGKHEEKQQEGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLIIEAPQVPYPYSTFGES
SEG      .....
PRD      ecccchhhhhccccceeeccccccccccccccceeeccccccceeecccccccccccccc

SEQ      SFNNELPQDSQEVCT
SEG      .....
PRD      cccccccccceeeccc

```

Prosites for DKFZphute1_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTYLASE ASP	PDOC00125

Pfam for DKFZphutel_23e13.3

HMM_NAME	Heat shock hsp20 proteins		
HMM	*AMMrpPVDWRE....DpDHFeVrMDMPGFKPEEIKVkvEDNNVlvIeG A P++ R + ++V++++ FKPEE+ VK+ D+ +++++G		
Query	77	ARFGVPAEGR-TPPFPFGPFWKVCNVVHVSFKPEELMVKTKDG-YVEVSG	123
HMM	EHEREEEREDdkWWWHERIYRHfMRRFrLPENVDpDqIkAsMsDngVLTi +HE E++ + + ++ F +++LP +VDP + AS+S++G+L I		
Query	124	KHE---EKQQ---EGGIVSKNfTKKIQLPAEVDpVTVFASLSPEGLLII	166
HMM	TVPKpEP* ++P ++P		
Query	167	EAPQVPP 173	

DKFZphut1_23g11

group: uterus derived

DKFZphut1_23g11 encodes a novel 256 amino acid protein with similarity to S.pombe SPAC31G5.12c and S. cerevisiae Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GCGGTCGCGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGGATGGG CCGGCAGGGC
151 CGGGAGTAAC GGGACGTCCG CGCGGAGCTT CTCCCCCGG ATACAGTGCG
201 GCCCCAGCGG AGGCCGCGGC GCCGCCCTCC GATCTTGAAG AGCCCGCGCT
251 GCGCGGAGCC CGCCCCCGCC TCGCACCCGG CACCGACGCG GAGCGACCAG
301 CCCAGCCAGA CCCGGCCCCG CGCGGCCCTG TCTAACCCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAAG ACATGAAGCT
401 ATTGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGCTACTC ATGTAAGATG
501 GCAGGAGACG ACAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACTTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAAGCCAA GCGGTCGAGG AGGAGGGCCC CCTCAGTGAC
651 AAGTGCAGCC GCAGAGCCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
751 GGGAGCCCCG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCTTC
951 AACTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCCCAGCT
1251 GCTGGCCAGA CCCTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCCTAGC CCTTTGGCTC CATCCTGTGG ATGCCCACTC ACCCCTCAGA
1351 CTCCTGCTGC CCATGCTGTG GCCGGACTTG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGCGACT GCCCTGCCCA AATGAAGTGC CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCCAA
1501 GGGTCTGTGG CCGGAGGCCC CACGAGCAGG CCCAGCAGT CACCGGCTCT
1551 GGTCTTGGGC CGGCCCGGCT GCCACCTGT ACCCCACCT CGCCATTG
1601 GCCGCTGCA CTGAGTGTC CTTTGCTGCA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256
Category: similarity to known protein


```

1 MKLLENSFE AINSQTLVET GDAHIIGRIE SYSCKMAGDD KHMFKQFCQE
51 GQPHVLEALS PPQTSGLSPLS RLSKSQGGEE EGPLSDKCSR KTLFYLIATL
101 NESFRPDYDF STARSHEFSR EPSLSWVVNA VNCSLFSAVR EDFKDLKPQL
151 WNAVDEEICL AECDIYSYNP DLSDPFGED GSLWSFNFF YNKRLLKRVF
201 FSCRSISGST YTPSEAGNEL DMELGEEVEE EESRSRGSGA EETSTMEEDR
251 VPVICI

```

BLASTP hits

Entry SPAC31G5_12 from database TREMBL:
 gene: "SPAC31G5.12c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c31G5.
 Score = 272, P = 9.3e-24, identities = 51/127, positives = 80/127

Entry SPD656_1 from database TREMBL:
 product: "ORF N150"; Yeast DNA for bfr2+ protein/pad1+ protein/sks1+
 protein, ORF N313, ORF N150, complete cds, and for ORF N118, partial
 cds.
 Score = 263, P = 8.4e-23, identities = 50/127, positives = 79/127

Entry S50986 from database PIR:
 MAF1 protein - yeast (Saccharomyces cerevisiae) >SWISSPROT:MAF1_YEAST
 MAF1 PROTEIN. >TREMBL:SC19492_1 gene: "MAF1"; product: "Maf1p";
 Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds.
 >TREMBL:SC8119_11 gene: "MAF1p"; product: "Maf1p"; S.cerevisiae
 chromosome IV cosmid 8119.
 Score = 180, P = 2.3e-17, identities = 43/133, positives = 75/133

Entry AF098499_2 from database TREMBL:
 gene: "C43H8.2"; Caenorhabditis elegans cosmid C43H8.
 Score = 263, P = 9.2e-23, identities = 78/252, positives = 118/252

Alert BLASTP hits for DKFZphut1_23g11, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_23g11, frame 3

Report for DKFZphut1_23g11.3

```

[LENGTH]      256
[MW]           28869.95
[PI]           4.51
[HOMOL]        TREMBL:SPAC31G5_12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR005c]
6e-13
[PROSITE]      MYRISTYL          3
[PROSITE]      CK2_PHOSPHO_SITE    5
[PROSITE]      PKC_PHOSPHO_SITE    6
[PROSITE]      ASN_GLYCOSYLATION   3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      7.81 %

```

```

SEQ  MKLLENSFEAINSQTLVETGDAHIIGRIESYSCKMAGDDKHMFKQFCQEGQPHVLEALS
SEG  .....
PRD  cccccchhhhhhhhhhhccccceeeccchhhhhccchhhhhhhhhccccceeeccc

SEQ  PPQTSGLSPLSRLSKSQGGEEEGPLSDKCSRKTLFYLIATLNESEFRPDYDFSTARSHEFSR
SEG  .....
PRD  cccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccc

SEQ  EPSLSWVVNAVNCSLFSVREDFKDLKPQLWNAVDEEICLAECDIYSYNPDLDSDPFGED
SEG  .....
PRD  cccccchhhhhhhhhhhhhchhhhhhhhhhhhhhhhhccccceeecccccccccccc

SEQ  GSLWSFNFFYNKRLLKRVFFSCRSISGSTYTPSEAGNELDMELGEEVEEESRSRGSGA
SEG  .....
PRD  cccccceeechhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhcccccc

SEQ  EETSTMEEDRVPVICI
SEG  xx.....
PRD  cccccccccceeeccc

```


Prosites for DKFZphut1_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_23g11.3)

DKFZphutel_24c19

group: transmembrane protein

DKFZphutel_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown

membrane regions: 1

Summary DKFZphutel_24c19 encodes a novel 195 amino acid protein, with no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp

Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
1  ACGAGTCAGC CAAAGATGGC TCGCCCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAAACAGT CTTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTTCT ATTTTGCTCC AGACTATGTT TTCAGCATAC CTGGGCTCTG
651 AACAAATATA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCCT GATTTTAAAC AAATATGTAA ACAAATAAAT AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195
Category: putative protein

```
1  MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51  NSLFRRILNV TKARIAAGLP MAGIPFLTTD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVERKML FPILLQTMFS AYLGSEQYKL LIKALQLSEP GKEIH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1_24c19, frame 2

Report for DKFZphut1_24c19.2

```

[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSLFRRILNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPMAGIPFLTTLTYRCFVSFPLNTGDLDCETCTITRSGTLGLVIGGLYPVF
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPILLQTMFSAYLGSEQYKL
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....

```

Prosite for DKFZphut1_24c19.2

```

PS00001      11->15  ASN_GLYCOSYLATION      PDOC00001
PS00001      34->38  ASN_GLYCOSYLATION      PDOC00001
PS00001      59->63  ASN_GLYCOSYLATION      PDOC00001
PS00005      18->21  PKC_PHOSPHO_SITE      PDOC00005
PS00005      82->85  PKC_PHOSPHO_SITE      PDOC00005
PS00005     151->154 PKC_PHOSPHO_SITE      PDOC00005
PS00006      13->17  CK2_PHOSPHO_SITE      PDOC00006
PS00008      40->46  MYRISTYL               PDOC00008
PS00008      47->53  MYRISTYL               PDOC00008
PS00008      68->74  MYRISTYL               PDOC00008
PS00008     110->116 MYRISTYL               PDOC00008
PS00008     127->133 MYRISTYL               PDOC00008
PS00008     142->148 MYRISTYL               PDOC00008

```

(No Pfam data available for DKFZphut1_24c19.2)

DKFZphut1_24e11

group: intracellular transport and trafficking

DKFZphut1_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits
potential start at 184,
TRANSMEMBRANE 4
function in the transport of nucleosides and/or nucleoside derivatives
between the cytosol and
the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp
Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
1  ACGCGTCCGG CAGAAGCTCG GAGCTCTCGG GGTATCGAGG AGGCAGGCCC
51  GCGGGCGCAC GGGCGAGCGG GCCGGGAGCC GGAGCGGCGG AGGAGCGCGC
101 AGCAGCGGCG CGGCGGGCTC CAGGCGAGGC GGTGACGCTC CCTGAAAACT
151 TGCGCGCGCG CTCGCGCCAC TCGCGCCGGA GCGATGAAGA TGGTCGCGCC
201 CTGGACGCGG TTCTACTCCA ACAGCTGCTG CTTGTGCTGC CATGTCCGCA
251 CCGGCACCAT CCTGCTCGGC GTCTGGTATC TGATCATCAA TGCTGTGGTA
301 CTGTTGATTT TATTGAGTGC CCTGGCTGAT CCGGATCAGT ATAACTTTTC
351 AAGTTCTGAA CTGGGAGGTG ACTTTGAGTT CATGGATGAT GCCAACATGT
401 GCATTGCCAT TCGGATTTCT CTTCATGA TCCTGATATG TGCTATGGCT
451 ACTTACGGAG CGTACAAGCA ACGCGCAGCC TGGATCATCC CATTCTCTTG
501 TTACCAGATC TTTGACTTTG CCCTGAACAT GTTGGTTGCA ATCACTGTGC
551 TTATTTATCC AAACCTCCAT CAGGAATACA TACGGCAACT GCCTCCTAAT
601 TTTCCCTACA GAGATGATGT CATGTCACTG AATCCTACCT GTTTGGTCCT
651 TATTATTCTT CTGTTTATTA GCATTATCTT GACTTTTAAG GGTTACTTGA
701 TTAGCTGTGT TTGGAAGTGC TACCGATACA TCAATGGTAG GAACTCCTCT
751 GATGTCCCTG TTTATGTTAC CAGCAATGAC ACTACGGTGC TGCTACCCCC
801 GTATGATGAT GCCACTGTGA ATGGTGCTGC CAAGGAGCCA CCGCCACCTT
851 ACGTGTCTGC CTAAGCCTTC AAGTGGGCGG AGCTGAGGGC AGCAGCTTGA
901 CTTTGCAGAC ATCTGAGCAA TAGTTCTGTT ATTTCACTTT TGCCATGAGC
951 CTCTCTGAGC TTGTTTGTG CTGAAATGCT ACTTTTAAAT ATTTAGATGT
1001 TAGATTGAAA ACTGTAGTTT TCAACATATG CTTTGCTAGA ACACTGTGAT
1051 AGATTAACTG TAGAATTCTT CCTGTACGAT TGGGGATATA ACGGGCTTCA
1101 CTAACCTTCC CTAGGCATTG AAACCTCCCC CAAATCTGAT GGACCTAGAA
1151 GTCTGCTTTT GTACCTGCTG GGCCCCAAAG TTGGGCATTT TTCTCTCTGT
1201 TCCCTCTCTT TTGAAAATGT AAAATAAAAC CAAAAATAGA CAACTTTTTC
1251 TTCAGCCATT CCAGCATAGA GAACAAAACC TTATGGAAC AGGAATGTCA
1301 ATTGTTGTAAT CATTGTTCTA ATTAGGTAAG TAGAAGTCCT TATGTATGTG
1351 TTTCAAGAAT TTCCCCCACA ACATCCTTTA TGACTGAAGT TCAATGACAG
1401 TTTGTGTTTG GTGGTAAAGG ATTTTCTCCA TGGCCTGAAT TAAGACCATT
1451 AGAAAGCACC AGGCCGTGGG AGCAGTGACC ATCTACTGAC TGTTCTTGTTG
1501 GATCTTGTGT CCAGGGACAT GGGGTGACAT GCCTCGTATG TGTTAGAGGG
1551 TGGAATGGAT GTGTTTGGCG CTGCATGGGA TCTGGTGCCC CTCTTCTCCT
1601 GGATTCACAT CCCCACCCAG GGCCCCGCTT TACTAAGTGT TCTGCCCTAG
1651 ATTGGTTCAA GGAGGTCATC CAACTGACTT TATCAAGTGG AATTGGGATA
1701 TATTTGATAT ACTTCTGCCT AACAAATGAG AAAAGGGTTT TCTTTTCCCT
1751 GCAAGCTACA TCCTACTGCT TTGAACCTCC AAGTATGCTC AGTCACCTTT
1801 TAAATGTATA ACATTTTCAG AAAAATGAGG ATTGCCTTCC TTGTATGCGC
1851 TTTTACCTTT GACTACCTGA ATTGCAAGGG ATTTTATAT ATTCAATATG
1901 TACAAAGTCA GCAACTCTCC TGTGGTTTCA TTATTGAATG TGCTGTAAT
1951 TAAGTCGTTT GCAATTAAAA CAAGTTTTCG CCACATCCAA AAAAAAAAAA
2001 AAAAA
```

BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.
Score = 1629, P = 3.1e-67, identities = 343/354

Medline entries

96199248:
Identification of a novel membrane transporter
associated with intracellular membranes by
phenotypic complementation in the yeast
Saccharomyces cerevisiae.

Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226
Category: strong similarity to known protein

```

1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP
51 DQYNFSSSEL GGDEFMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN
151 PTCVLIIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT
201 TVLLPPYDDA TVNGAAKEPP PPYVSA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24e11, frame 1

SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981_1 product: "E3 protein"; Human retinoic acid-inducible
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108).
Length = 233

HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53
Identities = 102/221 (46%), Positives = 148/221 (66%)

```

Query:      9 RFYSNCCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSELGGDF- 64
              RFYS  CC CCHVRTGTI+LG WY+++N ++ ++L  + P+  N  +G  +
Sbjct:     13 RFYSTRCCGCCHVRTGTIILGTWYMVVNLLMAILLTVEVTHPNMSPAVNIQYEVIGNYYS 72

Query:     65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123
              E M D N C+ A+S+LM +I +M YGA  + W+IPFFCY++PDF L+ LVAI+ L
Sbjct:     73 SERMAD-NACVLEAVSVLMFISSMLVYGAISYQVGWLIPFFCYRLFDFVLSCLVAISSL 131

Query:    124 IYPNSIQEYIRQLPPNFYRDDVMSVNPCTCLVLIILLFISIILTFKGYLISCVWNCYRYI 183
              Y  I+EY+ QLP +FPY+DD+++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI
Sbjct:    132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCWNCYKYI 190

Query:    184 NGRNSSDVLVYVTSN-DTTVLLPPYDDATVNGAAKEPPPPYVSA 226
              N RN  ++ VY  +LP Y+ A V  KEPPPPY+ A
Sbjct:    191 NNRNVPEIAVYPAFEAPPQYVLPYEMA-VKMPEKEPPPPYLP 233

```

Pendant information for DKFZphut1_24e11, frame 1

Report for DKFZphut1_24e11.1

[LENGTH] 226
[MW] 25419.11

535

DKFZphutel_24j6

group: cell structure and motility

DKFZphutes1_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CAR1).

The novel protein is very similar to Car1 and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CAR1 A.thaliana T19C21.5

complete cDNA, complete cds, EST hits
potential frame shift at Bp 1241 according to CAR1
but frame shift might be in CAR1 sequence!
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```

1  ACGCGTCCGA GCTGGCTCAG GGCCTCCGCT AGGCTCGGAC GACCTGCTGA
51 GCCTCCCAAA CCGCTTCCAT AAGGCTTTGC CTTTCCAAC TCAGCTACAG
101 TGTAGCTAA GTTGGAAAG AAGGAAAAA GAAATCCCT GGGCCCTTT
151 TCTTTTGTTC TTTGCCAAAG TCGTCGTTGT AGTCTTTTG CCCAAGGCTG
201 TTGTGTTTT AGAGGTGCTA TCTCCAGTTC CTTGCACTCC TGTTAAACAAG
251 CACCTCAGCG AGAGCAGCAG CAGCGATAGC AGCCGCAGAA GAGCCAGCGG
301 GGTCCGCTAG TGTATGACC AGGGCGGGAG ATCACAACCG CCAGAGAGGA
351 TGCTGTGGAT CCTTGGCCGA CTACCTGACC TCTGCAAAAT TCCTTCTCTA
401 CCTTGGTCAT TCTCTCTCTA CTTGGGGAGA TCGGATGTGG CACTTGGCGG
451 TGTCTGTGTT TCTGGTAGAG CTCTATGGAA ACAGCCTCCT TTTGACAGCA
501 GTCTACGGGG TGGTGGTGGC AGGGTCTGTT CTGGTCCTGG GAGCCATCAT
551 CGGTGACTGG GTGGACAAGA ATGCTAGACT TAAAGTGGCC CAGACCTCGC
601 TGGTGGTACA GAATGTTTCA GTATCCTGT GTGGAATCAT CCTGATGATG
651 GTTTCTTAC ATAAACATGA GCTTCTGACC ATGTACCATG GATGGGTTCT
701 CACTTCCCTG TATATCCTGA TCATCACTAT TGCAAAATATT GCAAATTTGG
751 CCAGTACTGC TACTGCAATC ACAATCCAAA GGGATTGGAT TGTTGTTGTT
801 GCAGGAGAAG ACAGAAGCAA ACTAGCAAA ATGAATGCCA CAATACGAAG
851 GATTGACCG TTAACCAACA TCTTAGCCCC CATGGCTGTT GGCCAGATTA
901 TGACATTGG CTCCCCAGTC ATCGGCTGTG GCTTTATTTC GGGATGGAAC
951 TTGGTATCCA TGTGCGTGGA GTACGTCCTG CTCTGGAAGG TTTACCAGAA
1001 AACCCCAAG CTAGCTGTGA AAGCTGGTCT TAAAGAAGAG GAAACTGAAT
1051 TGAACAGCT GAATTACAC AAGATACTG AGCCAAAACC CCTGGAGGGA
1101 ACTCATCTAA TGGGTGTGAA AGACTCTAAC ATCCATGAGC TGAACATGA
1151 GCAAGAGCCT ACTTGTGCCT CCCAGATGGC TGAGCCCTTC CGTACCTTCC
1201 GAGATGGATG GGTCTCCTAC TACAACCAGC CTGTGTTTCT GGCTGGCATG
1251 GGTCTTGCTT TCCTTTATAT GACTGTCTTG GGCTTTGACT GCATCACCAC
1301 AGGGTACGCC TACACTCAGG GACTGAGTGG TTCCATCCTC AGTATTTTGA
1351 TGGGAGCATC AGCTATAACT GGAATAATGG GAACTGTAGC TTTTACTTGG
1401 CTAGTCTGAA AATGTGGTTT GGTTCGGACA GGTCTGATCT CAGGATTGGC
1451 ACAGCTTTCC TGTTTGATCT TGTGTGTGAT CTCTGTATTC ATGCCCTGGA
1501 GCCCCCTGGA CTTGTCCGTT TCTCCTTTTG AAGATATCCG ATCAAGGTTT
1551 ATTCAGGAG AGTCAATTAC ACCTACCAAG ATACCTGAAA TTACAACTGA
1601 AATATACATG TCTAATGGGT CTAATTCTGC TAATATTGTC CCGGAGACAA
1651 GTCCCTGAATC TGTGCCATA ATCTCTGTCA GTCTGCTGTT TGCAGGCGTC
1701 ATTGCTGCTA GAATCGGTCT TTGGTCCTTT GATTAACTG TGACACAGTT
1751 GCTCAAGAA AATGTAATTG AATCTGAAAG AGGCATTATA AATGGTGATC
1801 AGAACTCCAT GAACTATCTT CTTGATCTTC TGCATTTTCA CATGGTCATC
1851 CTGGCTCCAA ATCCTGAAGC TTTTGGCTTG CTCGTATTGA TTTCACTCTC
1901 CTTTGTGGCA ATGGGCCACA TTATGTATTT CCGATTGGCC CAAAATACTC
1951 TGGGAAACAA GCTCTTTGCT TCGGTCCTG ATGCAAAAGA AGTTAGGAAG
2001 GAAATCAAG CAAATACATC TGTGTTTGA GACAGTTTAA CTGTTGCTAT
2051 CCTGTACTA GATTATATAG AGCACATGTG CTTATTTTGT ACTGCAGAAT
2101 TCCAATAAAT GGCTGGGTGT TTTGCTCTGT TTTTACCACA GCTGTGCCCT
2151 GAGAACTAAA AGCTGTTTAG GAAACCTAAG TCAGCAGAAA TTAAGTATT
2201 AATTTCCTT ATGTTGAGGC ATGGAAAAA AATTGGAAAA GAAAACTCA
2251 GTTTAAATAC GGAGACTATA ATGATAACAC TGAATTCCTC TATTCTCAT
2301 GAGTAGATAC AATCTTACGT AAAAGAGTGG TTAGTCACGT GAATTCAGTT
2351 ATCATTTGAC AGATTCTTAT CTGTACTAGA ATTCAGATAT GTCAGTTTTC
2401 TGCAAAACTC ACTCTTGTTT AAGACTAGCT AATTTATTTT TTTGCTCTT
2451 AGTTATTTTT AAAACAAAT TCTTCAAGTA TGAAGACTAA ATTTTGATAA
2501 CTAATATTAT CCTTATTGAT CCTATTGATC TTAAGGTATT TACATGTATG

```



```

2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAGCACTT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAAGTA GATTATCAC ACAATGACTG CATACAGACT
2751 TCAAAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACATCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCCACA CTTGTGAAGG TTTTGTTTTA CAAATCACTT GATTTAACAC
2951 ACTCAGGTAG AATATTTTTA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAAGTC AGTTTGCAAC ATGTCTGTAC CAAGATGGTA CTTTGCCTTA
3151 ACCGTTTATA TGCACTTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TTCTTTATCC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATTG GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAATAA AAA

```

BLAST Results

Entry HS389210 from database EMBL:
human STS SHGC-10164.
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:
human STS WI-16551.
Score = 1193, P = 5.7e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571
Category: strong similarity to known protein

```

1 MTRAGDHNRO RGCCGSLADY LTSAKFLLYL GHSLSTWGDR MWHFAVSVFL
51 VELYGNLALL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKHEL LTMVHGWLVT SCYILIITIA NIANLASTAT
151 AITTIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNIALPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCVEY VLLWKVYQKT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTHLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDGWV
301 SYYNQPVFLA GMGLAFLYMT VLGFOCITTG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAF TWLRRKGLV RTGLISGLAQ LSCLILCVIS VFMPGSPDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEI YMSNGSNSAN IVPETSPESV
451 PIISVSLLEA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24j6, frame 3

TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P
= 2.8e-60

TREMBL:AF039046_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.
Length = 405

HSPs:

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151
Identities = 288/319 (90%), Positives = 297/319 (93%)

```

Query:   1 MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGD RMWHFAVSVFLVELYGNLSLL 60
          MT++ D  Q GCCGSLA+YLTSAKFLLYLGHSLSTWGD RMWHFAVSVFLVELYGNLSLL
Sbjct:   1 MTKSRDQTHQEGCCGLANYLTSAKFLLYLGHSLSTWGD RMWHFAVSVFLVELYGNLSLL 60

Query:  61 TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL 120
          TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHK+EL
Sbjct:  61 TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKNEL 120

Query: 121 LTMHGWVLTSCYILIITIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI 180
          L MYHGWVLT CYILIITIANIANLASTATAITIQRDWIVVAGE+RS+LA+MNATIRRI
Sbjct: 121 LNMHGWVLTVCYILIITIANIANLASTATAITIQRDWIVVAGENRSRLADMNATIRRI 180

Query: 181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK 240
          DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEY LLWKVYQKTPALAVKA LK
Sbjct: 181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAALK 240

Query: 241 EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVW 300
          EE+ELKQL KDTEPKPLEGTHLMG KDSNI ELE EQEPTCASQ+AEFPRTFRDGVW
Sbjct: 241 VEESELKQLTSPKDTEPKPLEGTHLMGEKDSNIRELECEQEPTCASQIAEPFRTFRDGVW 300

Query: 301 SYYNQPVFLAGMGLAF-LY 318
          SYYNQPVFL G F LY
Sbjct: 301 SYYNQPVFLGWHGPGFPLY 319

```

Pedant information for DKFZphutel_24j6, frame 3

Report for DKFZphutel_24j6.3

```

[LENGTH]      571
[MW]           62542.72
[pI]           6.08
[HOMOL]        TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator (CAR1) mRNA, complete cds. 1e-141
[BLOCKS]       BL00341D
[PROSITE]      MYRISTYL 15
[PROSITE]      MITOCH_CARRIER 1
[PROSITE]      CK2_PHOSPHO_SITE 6
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      PKC_PHOSPHO_SITE 4
[PROSITE]      ASN_GLYCOSYLATION 4
[PFAM]         Laminin B (Domain IV)
[KW]           TRANSMEMBRANE 4
[KW]           LOW_COMPLEXITY 8.76 %

```

```

SEQ  MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGD RMWHFAVSVFLVELYGNLSLL
SEG  .....
PRD  cccccccccccccccccchhhhhhhheeeccceccchhhhhhhheeecccccce
MEM  .....MMMMMMMMMMMM

```

```

SEQ  TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ehhhhhhhccceeecccccchhhhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

```

SEQ  LTMHGWVLTSCYILIITIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhccccchhhhhhhhhhhhhhhhhhhhhheeeccceeecccccchhhhhhhhhhhhh
MEM  MMMMMMMM.....

```

```

SEQ  DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG  .....
PRD  hhhhhhccceeeccceeeccceeeccchhhhhhhhhhhhhccchhhhhhhhhhh
MEM  .....

```

```

SEQ  EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVW
SEG  .....
PRD  hhhhhhhhhcccccceeecccccceeecccccceeecccccceeecccccceee
MEM  .....

```

```

SEQ  SYYNQPVFLAGMGLAFLYMTVLGFDCTTGAYTQGLSGSILSILMGASAITGIMGTVA
SEG  .....
PRD  eeccceeeccchhhhhcccccceeecccccceeecccccceeecccccceeehhhhhh

```


Prosites for DKFZphutel 24j6.3

Pfam for DKFZphute1 24j6.3

539

DKFZphutel_2h3

group: differentiation/development

DKFZphutel_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits
complete cds according to E25 start at Bp 56
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp
Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```
1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GCGGACGCG CGGGCCGGCG
51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
101 GACAAGGCTG ACAAGGCGTC GCGCTCGGCC CCTGCGCCGG CCTCGGCCAC
151 CGAGATCCTG CTGACGCGCG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
201 CCAAGAGGGG GAGCTCAGTG GCGGCGGTGT GCTACCTGTC GATGGGCATG
251 GTCGTGCTGC TCATGGGCCT CGTGTTCGCC TCTGTCTACA TCTACAGATA
301 CTTCTTTTCT GCACAGCTGG CCGAGATAAA CTTCTCCGC TGTGGTGTGC
351 TGTATGAGGA CTCCCTGTCC TCCCAGGTCC GGAATCAGAT GGAGCTGGAA
401 GAGGATGTGA AAATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
451 TGTGCCCCAG TTTGGCGGCG GTGACCCTGC AGACATCATC CATGACTTCC
501 AGCGGGGTCT GACTGCGTAC CATGATATCT CCCTGGACAA GTGCTATGTC
551 ATCGAATCTA ACACCACCAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
601 CCTCATGAAC GTGAAGAGGG GGACCTACCT GCCGACAGC TACATCATCC
651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAGGA GGCCCTGGGG
701 TCCTTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGGCCAAG AACTGCAATG
801 CCATCCGCCA CTTGAGAAC ACCTTCGTGG TGGAGACGCT CATCTGCGGG
851 GTGGTGTGAG GCCCTCCTCC CCCAGAACCC CCTGCCGTGT TCCTCTTTTC
901 TTCTTTCCAG CTGCTCTCTG GCCCTCCTCC TTCCCCCTGC TTAGCTTGTA
951 CTTTGACGCG GTTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CCTGCCACCC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 CTCTGCTGAC CTGGGTGTGG CGGAGGGAGA GCGATGCTG CAAAGTGTTT
1101 TCTGTGTCCC ACTGTCTTGA AGCTGGGCTT GCCAAAGCCT GGGCCACAG
1151 CTGCACCGGC AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGGCATGTG
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GCGGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCCTC AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCCTAGA CCTGGGCTT GAGCTGAGGA
1351 TGGGACGATG TGCCACAGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCAGGG ACTCTGTGAG TGCCTTCAGC CCACCAGCAG
1451 GAGCTTGGAG TTTGGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAAGAAGC AAGGAGCTAG GACCCCCAGT CCTGCCCCC
1551 AGGAGCACAA GCAGGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGG AGGCAAGGTC ACTGCTCAGT CACGTCCACG GGGGACGAGC
1651 CGTGGGTCTT GCTGAGTAGG TGGAGCTCAT TGCTTCTCC AAGCTTGGAA
1701 CTGTTTGTAA AGATAACACA GAGGGAAAGG GAGAGCCACC TGGTACTTGT
1751 CCACCTGTCC TCCTCTGTTC TGAATTTCCA TCCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTTCCT TCTCACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGCTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAAATC TTAATGCTT TGTATATTTT
1951 CTCAATTAGA TCTCTTTTCA GAAGTGCTA TAGAACAATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGGCGG CCG
```

BLAST Results

Entry B64417 from database EMBL:
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.
Length = 715
Plus Strand HSPs:

Score = 1546 (232.0 bits), Expect = 7.8e-64, P = 7.8e-64
Identities = 310/311 (99%)

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA
library subtraction.
Molecular cloning and characterization of a gene belonging to a novel
multigene family of
integral membrane proteins.

Peptide information for frame 2

ORF from 56 bp to 856 bp; peptide length: 267
Category: strong similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGSSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VPVPQFGGSD PADIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFW ELLMNVKRGY YLPQTYIIQE
201 EMVUTEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_2h3, frame 2

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16).., N = 1, Score = 573, P = 1.3e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN).., N =
1, Score = 560, P = 3.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN).., N = 1,
Score = 456, P = 3.3e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16).
Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.3e-55, P = 1.3e-55
Identities = 117/264 (44%), Positives = 172/264 (65%)

```

Query: 1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY 60
      MVK+SF A+A + A+K ++ ++L+ P ++P G
Sbjct: 1 MVKVSFNSALA--HKEAANKEENS-----QVLILPPDAKEPEDVVVPAGHKRAWWC 51

Query: 61 LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL-----SQVRTQM-- 112
      + G+ +L G++ Y+Y+YF Q + CG+ Y ED LS +Q+++
Sbjct: 52 MCFGLAFMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKIEDGLSLPESGAQLKSARYH 108

Query: 113 ELEDVKIYLDENYERINVPVPQFGGSDPADIHDFQRLTAYHDISLDKCYVIELNTTI 172
      +E++++I +E+ E I+VPVP+P DPADI+HDF R LTAY D+SLDKCYVI LNT++
Sbjct: 109 TIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRLTAYLDLSLDKCYVIPLNTSV 168

Query: 173 VLPFRNFWEELLMNVKRGTYLPQTYIIQEEMVUTEHVSDKEALGSFIYHLCNGKDTYRLR 232
      V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+R
Sbjct: 169 VMPPKNFELLINIKAGTYLPQSYLIHEQMIQTDRIENVDQLGFFIYRLCRGKETYKLQR 228

Query: 233 RATRRRINKRGAKNCNAIRHFENTFVETLIC 264
      + + I KR A NC IRHFEN F +ETLIC
Sbjct: 229 KEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

```

Pedant information for DKF2phut1_2h3, frame 2

```

[LENGTH]      267
[MW]           30253.96
[pI]           8.16
[HOMOL]        SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
le-49
[PROSITE]      MYRISTYL      4
[PROSITE]      PRENYLATION   1
[PROSITE]      CAMP_PHOSPHO_SITE      3
[PROSITE]      CK2_PHOSPHO_SITE      3
[PROSITE]      TYR_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      4
[PROSITE]      ASN_GLYCOSYLATION      1
[KW]           TRANSMEMBRANE 1
[KW]           LOW COMPLEXITY      15.36 %

```

```

SEQ      MVKISFQPAVAGIKGDKADKASAPAPASATEILTPAREEQPPQHRSKRKSSVGGVCY
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      cccccccccchhhhhhhhhhhhhhhhhccccccceccccccccccccccccccccchh
MEM      .....MMM.....

SEQ      LSMGMVLLMGLVFASVYIYRFFLAQLARDNFFRCGVLIEDSLSSQVRTQMELEEDVKI
SEG      .XXXXXXXXXXXXX.
PRD      hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      YLDENYERINVPVPQFGGGDPADIHDFQRGLTAYHDISLQKCYVIELNTTIVLPPRNF
SEG      .....
PRD      hhccccceccccccccccccchhhhhhhhhhhhhhhccccceccccceccccchh
MEM      .....

SEQ      ELLMNVKRGTYLPQTYIIQEEMVVEHVSDEKALGSFIYHLCNGKDYRLRRRATRRRIN
SEG      .....XXXXXXXXXXXXX.....
PRD      hhhhhccccccccceeeehhhhhhhccccchhhhhheccccchhhhhhhhhhhhh
MEM      .....

SEQ      KRGAKNCNAIRHFENTFVVETLICGVV
SEG      xx.....
PRD      hhhhhccccceccccchhhhhheeeccc
MEM      .....

```

PS000001	169->173	ASN_GLYCOSYLATION	PDOC000001
PS000004	50->54	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	187->191	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	232->236	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	49->52	PKC_PHOSPHO_SITE	PDOC000005
PS000005	209->212	PKC_PHOSPHO_SITE	PDOC000005
PS000005	227->230	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000006	30->34	CK2_PHOSPHO_SITE	PDOC000006
PS000006	110->114	CK2_PHOSPHO_SITE	PDOC000006
PS000006	209->213	CK2_PHOSPHO_SITE	PDOC000006
PS000007	119->127	TYR_PHOSPHO_SITE	PDOC000007
PS000008	52->58	MYRISTYL	PDOC000008
PS000008	71->77	MYRISTYL	PDOC000008
PS000008	138->144	MYRISTYL	PDOC000008
PS000008	243->249	MYRISTYL	PDOC000008
PS002294	264->268	PRENYLATION	PDOC002266

542

DKFZphmcf1_lal1

group: transmembrane protein

DKFZphmcf1_lal1 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcf1_lal1 encodes a novel 393 amino acid protein, with similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits
potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGGCCCCAG CCCCGGAAGA GCCGCCTCAG CCGGGGGGAG TTGCTCGGAC
51  TCAAACGTCC AGTCCTCGTG CGACCGCGCT GGGTCGGAAG TGAGCAGGCT
101 GAGGCCACCA TGGAGCAGTG TGCGTGCGTG GAGAGAGAGC TGGACAAGGT
151 CCTGCAGAAG TTCCTGACCT ACGGGCAGCA CTGTGAGCGG AGCCTGGAGG
201 AGCTGCTGCA CTACGTGGGC CAGCTGCGGG CTGAGCTGGC CAGCGCAGCC
251 CTCCAGGGGA CCCCTCTCTC AGCCACCCTC TCTCTGGTGA TGTCACAGTG
301 CTGCCCGAAG ATCAAAGATA CGGTGCAGAA ACTGGCTTCG GACCATAAGG
351 ACATTACAGC CAGTGTATCC CGAGTGGGCA AAGCCATTGA CAGGAACCTC
401 GACTCTGAGA TCTGTGGTGT TGTGTGAGAT GCGGTGTGGG ACGCGCGGGA
451 ACAGCAGCAG CAGATCCTGC AGATGGCCAT CGTGGAAACAC CTGTATCAGC
501 AGGGCATGCT CAGCGTGGCC GAGGAGCTGT GCCAGGAATC AACGCTGAAT
551 GTGGACTTGG ATTTCAAGCA GCCTTTCCTA GAGTTGAATC GAATCCTGGA
601 AGCCCTGCAC GAACAAGACC TGGGTCTGCG GTTGGAAATGG GCCGTCTCCC
651 ACAGGCAGCG CCTGCTGGAA CTCACAGCT CCCTGGAGTT CAAGCTGCAC
701 CGACTGCACT TCATCCGCCT CTTGGCAGGA GGCCCCGCGA AGCAGCTGGA
751 GGCCCTCAGC TATGCTCGGC ACTTCCAGCC CTTTGTCTCG CTGCACCAGC
801 GGGAGATCCA GGTGATGATG GGCAGCCTGG TGTACCTGCG GCTGGGCTTG
851 GAGAAGTCAC CCTACTGCCA CCTGCTGGAC AGCAGCCACT GGGCAGAGAT
901 CTGTGAGACC TTTACCCGGG ACGCCTGTTC CCTGCTGGGG CTTTCTGTGG
951 AGTCCCCCTC TAGCGTCAGC TTTGCCTCTG GCTGTGTGGC GCTGCTGTG
1001 TTGATGAACA TCAAGGCTGT GATTGAGCAG CGGCAGTGCA CTGGGGTCTG
1051 GAATCACAAG GACGAGTTAC CGATTGAGAT TGAAGTAGGC ATGAAGTGCT
1101 GGTACCACTC CGTGTTCGCT TGCCCCATCC TCCGCCAGCA GACGTCAGAT
1151 TCCAACCTC CCATCAAGCT CATCTGTGGC CATGTTATCT CCCGAGATGC
1201 ACTCAATAAG CTCATTAATG GAGGAAAGCT GAAGTGTCCC TACTGTCCCA
1251 TGGAGCAGAA CCCGGCAGAT GGGAAACGCA TCATATTCTG ATTCTACCT
1301 GGAAGGAATT TTGTTGAAAG GGGTTTTCAC CTGTGAGCCT TGGTCTGTCT
1351 CGGTAGGGTG GTCAACTTCA GTGGACTGTG GTTGGTTTCA GAGCGCCTGG
1401 CTGAGGAGTT CCACTGAGGG GAGCACTGGA GCAGCCCTTT GGCAGAGGCT
1451 GAGGAGGGAG ATGGACCAGC CCACGCCTGG CACCTGGCTC CATGGCATAA
1501 GGAAAGGGAG ATGCTGGCCT CTGTGCTCCT GCTGTCTTTT CCTGTTCTG
1551 TTTGCGTTTG ACTTAGTAGC AACCGACAGA GTGGCAAGGG ATTTGGTCTT
1601 CAGCAGTAGA CATCCTTCCA CCCCTGCCCT CAGCCAAGTC TCTTGCTGCC
1651 ATGCCAATGC TATGTCCACC CTTGCCCTC GGCCCAAGAG TGTCACGCGG
1701 TGGCCCACTT CTTCTCCCA CTACAGCCTC AACAGTATGT ACCATCTCCC
1751 ACTGTAAATA GTCCAGTTA GAACGGAATG CCGTTGTTTT ATAACTTTGA
1801 ACAAATGTAA AAAAAAAAAA

```

BLAST Results

Entry HS579359 from database EMBL:

human STS WI-6350.

Score = 1027, P = 9.9e-40, identities = 207/209

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQHCERSLE ELLHYVGQLR AELASAAALQG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDSQ
101 ICGVVSDAVW DAREQQQQIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETFTF DACSLLGLSV ESPLSVSFAS GCVALPVLNM
301 IKAVIEQRQC TGVWNHKDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLCIGHVI SRDALNKLIN GGLKCPYCP MEQNPADGKR IIF
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1a11, frame 2

TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3.
 Length = 398

HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 55/142 (38%), Positives = 89/142 (62%)

```

Query: 252 YCHLLDSSHWAEICETFTFDACSLLGLSVESPLSVSFASGCVALPVLNMNIKAVIEQRQCT 311
      Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ ++++++
Sbjct: 258 YIDVLDLD-WKSLELLFVREFCAALGMSLESPLDIVNAGAIALPILLKMSIMKKKHT 316

Query: 312 GVVWNHKDELPIEIELGMKCWYHSVFACPILRQQTSDSNPPKLCIGHVISRDALNKLING 371
      W + ELP+EI L +HSVF CP+ ++Q ++ NPP+ + CGHVI ++L +L
Sbjct: 317 --WTSQGELPVEIFLPSSYHFHSVFTCPVSKEQATEENPPMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKCPYCPMEQNPADGKRIIF 393
      G + KCPYCP E AD R+ F
Sbjct: 375 GSQRFKCPYCPNENVAADAIRVYF 398
  
```

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 51/221 (23%), Positives = 102/221 (46%)

```

Query: 22 GQHCERSLEELLHYVGQLRAELASAAALQGTPLSATLSLVMSQCCRKIKD TVQKLASDHKD 81
      G C L EL + + + L+ P ++ LV C K + L K
Sbjct: 15 GNKCLAKLNEL---ESILKDAKSKCLKD-PTTSMKELVA--CSEKTQQVFDDLKRTEKK 67

Query: 82 IHSSVSRVGKAIDRNFDSQICGVVSDAVWDAREQQQQILQMAIVEHLYQQGMLSVAEELC 141
      H+S++R GK +++ F+ ++ + +++++++ + A+ H ++QG + +A C
Sbjct: 68 FHTSLNRFGKTLEKKNFEDLEIKLHSSFESKKRE---IDTALS LHFFRQGDVELAHLFC 124

Query: 142 QESTLNVDLDFKQPFLELNRIEALHEQDLGPALEWAVSHRQRLLELNS SLEFKLHRLHF 201
      +E+ + + F L I++ + ++DL +EWA R L SSLE+ L +
Sbjct: 125 KEAGIEEPSSESLHVFTLLKSIQGIRDKDLKPIEWASQCRGYLERKGS SLEYTLQKYRL 184

Query: 202 IRLLAGGPAKQL-EALSYAR-HFQPFARLHQREIQVMGSLVY 242
      + K + A+ Y R + F + H +IQ M +L +
  
```


Pedant information for DKFZphmcfl_1a11, frame 2

Report for DKFZphmcf1_1a11.2

```
SEQ      MEQCACVERELDKVLQKFLTGYQHCHERSLEELLHYVQQLRAELASAAOQTPLSATLSLV  
PRD      cccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhhh  
MEM      .....  
  
SEQ      MSQCCRKIKDTVQKLASHKHDIHSSVSVRVGKAIDRNFDSIEICGVVSDAVWDAREQQQQI  
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeechhhhhhhhhhhhhhh  
MEM      .....  
  
SEQ      QMAIVEHLYQQGMLSAEELCQUESTLNVLDLFKQPFLELNRIREALHEQDLGPALAWVS  
PRD      hhhhhhhhhhhhhccccchhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhccccchhhhhhh  
MEM      .....  
  
SEQ      HRQRLELNSSLEFKLRHLHFIRLLAGGPAKQLEALSARHFQPFARLHQREIQVMMGSL  
PRD      hhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
MEM      .....  
  
SEQ      VYLRLGLEKSPYCHLLDSSHWAEICETFTDRACSLLGLSVESPLSVSFASGCVALPVLMM  
PRD      hhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM  
  
SEQ      IKAVIEQRQCTGVWNHKDELPIEIELGMKCWYHSVFACPILRQQTSDSNPPIKLICGHVI  
PRD      hhhhhhhhhhhhhccccccccccccceeeecceeeeeeeecchhhhhhhcccccccccccccccccee  
MEM      MMMMMM.....  
  
SEQ      SRDALNKLINGGKLKPYPCFMEQNPDGKRIF  
PRD      eehhhhhhhhhccccccccccccccccchhhhhhhccc  
MEM
```

Prosites for DKFZphmcf1_1a11.2

PS000001	189->193	ASN_GLYCOSYLATION	PDOC000001
PS000005	180->183	PKC_PHOSPHO_SITE	PDOC000005
PS000006	28->32	CK2_PHOSPHO_SITE	PDOC000006
PS000006	135->139	CK2_PHOSPHO_SITE	PDOC000006
PS000006	190->194	CK2_PHOSPHO_SITE	PDOC000006
PS000007	211->219	TYR_PHOSPHO_SITE	PDOC000007
PS000007	27->36	TYR_PHOSPHO_SITE	PDOC000007
PS000007	244->253	TYR_PHOSPHO_SITE	PDOC000007
PS000008	37->43	MYRISTYL	PDOC000008
PS000008	50->56	MYRISTYL	PDOC000008
PS000009	387->391	AMIDATION	PDOC000009
PS000013	282->293	PROKAR_LIPOPROTEIN	PDOC000013

545

DKFZphmcfl_1c23

group: mammary carcinoma derived

DKFZphmcfl_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
1  AACTGGCCCC CTCCCCACC CCCTGCCCT GAGGAGCAGG ACCTGTCCAT
51  GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCCTTCAGGC TCCCAGAGC TTGTCAGCTC CCCGGCTGCT
151 TCGTCTCTCT CAGCTACTGC TTTGCAGATT CAGCCCCCGG GTAGCCCAGA
201 CCCTCTCTCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC CCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCCTCCCA GGGAGGACGT AGGTGCCGCC CTGGTCACGC CCTCGCTCCT
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCCACCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
451 GGGCGGGGCA GCCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCTGAG GCAGAGCCAC GGCCCTCCCA GTCCCTGCC
601 TCAACGGCCA GTTTCATCTT CTCCAAGGCG TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCGTG TCCCTTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAAGTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGCG CCCAAGAAAG
751 TCACCTAAGG CTCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
801 CCCACCCGCC TCCCCAGTT ACCCTCGAGC TGAGCCCTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAAATGGAG GTGTCTTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCCT
951 CCCGGGCTCA GACTCACAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGAGCAGCT
1051 CCAAGGACGA GAGGATACAG CAGACACAAC CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCCT TCGATACTTA TGCAAGCCTG GTGTTGCTCC
1151 TGTCTCTAGA GTCATCTGCG GCTCATGCCT TTTCCCGAAT GGGTTCACCT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCCCTG CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTCCTCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTGAAGAA GTGGCTGCAT
1401 GGCATCTCTC CAGGGGCCAG GAAGTTGCAT TCCAAGGGCC TGTTCATATG
1451 GCAGCAGAAT CCATCCCCCG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCCTCTGCG CCAGTCCAGT TTAATCTCTT TGGTTCCTGA AGGTGGCCAA
1551 GTCAATTGTT TCCACAGGCG TTCTTAGGCG TGGGGGCAGG TGTGGGGCTG
1601 TGGAAATTCCA AAGCACAAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACTCACCA ACCACCTTCC TGCCCTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCCTCCTT
1801 GACCACACTG GGTGGGTGGG GCCCCAGCA CTTCAGAGGC AGGAGCCTTT
1851 GGGCTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCCT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCTCA TTTCCGTGCA TGTCCTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGAAGGCCAC AGGTGCATTT TACTCAAGTC TTCTTAGTTC AATGAGGGGC
2101 ACCCAGTGCT TCTAGGGCAG GCTGGGTGGT GGTCCCCTAG GTATCAGCCT
2151 CTCTTACTGT ACTCTCCGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CACCTGTCTA GGCAAGCTGG CTTCCCCATT GGCCCTGTG GGTCCACAGC
2251 AGCGTGGCTG CCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTTCTCT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAAGTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCCACTG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTTCTGCTC
2601 AGAGCTCCTG ATCTACCCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTTCCC TCTCCCTCCT CCAACATTAC TGGAACTCTA
```



```

2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AAGTATATAT ATACATATAT ATATTCTTTT
2951 AAATTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAACTGTG TTTCAATTAA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311
 Category: putative protein
 Classification: unset

```

1 MADFPPPEEA FFSVASPEPA GPSGSPELVS SPAASSSSAT ALQIQPPGSP
51 DPPFPAPPAPA PASSAPGHVA KLPQKEPVGC SKGGGPPRED VGAPLVTPSL
101 LQMVRLRSVG APGGAPTAL GPSAPQKPLR RALSGRASPV PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQFN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERPVSPETQA DLQRNLVAEL RSISEQRPPQ APKKSPPKAPP PVARKPSVG
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRK RELAENGGLV QLVGPPEKMG
301 LPGSDSQREL A

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize
 Length = 1,188

HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15
 Identities = 81/269 (30%), Positives = 115/269 (42%)

```

Query:      5 PPPEEAFFS----VASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
           PPP      S      V SP P P SP      PA +SS ++ PP +P PPP +
Sbjct:    598 PPPPAPVASPPPPVKSPPPPTFVASPP---PPAPVASSPPPMKSPPPPTFVSSPPFPEKS 654

Query:     56 PPAPAPASSAPGHVAKLPQKEPVGC SKGGGPPREDVGAPLVTPSL LQMVRLRSVGAPGGA 115
           PP P PA S P + P P K PP ++ P + PS + P
Sbjct:    655 PPPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP 711

Query:     116 PTPALGPSAPQKPLRRA-LSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
           P+ PS P++P+ + ++SP PAP S +LA S + + PP
Sbjct:    712 PSSPEKPSPPKEPVSSPPQTPKSSPPAPVSSPPPTFVSSPPALAPVSSPPSVKSSPPPA 771

Query:     175 AEPRPPQSPASTASFIFSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPPQAPK 233
           PP +P +S +Q+ P +P++ L V+ + + PP AP
Sbjct:    772 PLSSPPAPQVKSS-----PPPVQVSSPPAPKSSPPLAP--VSSPPQVEKTSPPPA 823

Query:     234 KSPKAPPPVARKPSVG--PPPASPSYPRAEPLTAPPTNGLP 273
           SP P + P V V PPP SP P++PP P
Sbjct:    824 SSPLAPK-SSPHVVVSSPPPVKSSPPAPVSSPPLTPKP 864

```

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPELVSSPAASS---SSATALQIQPPGSPDPPPPAP---PAPAPASSAPGHV 69
 P P G P SP + PAAS+ S T + P P+P P P P P P +P
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSPTTPDVSPEPLPEFSPVPAPAPMPMTTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPEGGAPTALGFSAPQKP 128
 +P PV G S P V P + +V+L AP G+P P + ++P P
 Sbjct: 469 DYVPPTPPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188
 + G SP P P S + +K+ A G + P PPE P PP AS
 Sbjct: 529 I-----GSPSP-PPVSVVSPPPVKSPPPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSKGSRKLQLERPVSPTQADLQRNLVAELRSISEQRPPQAPKKSPPKAPPPVARKPS- 247
 + S L P P ++ VA + PP P SP P PVA P
 Sbjct: 578 PVKSPPPTLVASPP--PPVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVASSPPP 635

Query: 248 VGVPPP----ASPSYPRAEPLTAPPTNGLPHQTQD 277
 + PPP +SP P P PP P ++
 Sbjct: 636 MKSPPPTPVSSPPPEKSPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPELV--SSP--AASSSSATALQIQPPGSP-DPPPPAPPAPAPASSAPGHVA 70
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+
 Sbjct: 817 SPPPA-PLSSPPLAPKSSPPHHVVSSPPPVVKSPPPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ----KEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPEGGAPTALGFSAPQ 126
 P+ P + PP E +P TP L ++S P +P + P +
 Sbjct: 873 SPPEVVKPSTPPAPTTPVISPPSEPKSSPPPTPVSLPPPIVKSSPPAMVSSPMTKSSP 932

Query: 127 KPLRRAL---SGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P
 Sbjct: 933 PPVVVSSPPPTVKSSPPAPVSSPPATP--KSSPPAPVNL----P--PPEVKSSPPPTP 984

Query: 184 ASTASFIFSKGSRKLQLERPVSPTQADLQRNLVAELRSISEQRPPQAPKKSPPKAPPPVA 243
 S+ + P PE ++ V+ + PP AP SP PPPV
 Sbjct: 985 VSSPPPAKKSPPPPAPMSSPPPEVKSSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPVK 1042

Query: 244 RKPS---VGVPPPASPSPYRAEPLTAPP 268
 P V PPP S P P+++PP
 Sbjct: 1043 SPPPPAPVSSPPPPVKSPPPPAPISSPP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPELVSSPAAS-SSSATALQIQPPGSPDPPPPAPPAPAPAS 63
 PPP S PE + P P + P + T+++ PP PP P+P
 Sbjct: 639 PPPPTPVSSPPPEKSPPPPPAKSTPPPEEYTPPTSVKSSPPPEKSLPPPTLIPSPPP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPEGGAPTALGFS 123
 P K P K PP+E V +P TP V +P PTP P
 Sbjct: 699 QEKPPTPSTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPAPVSSP--PPTPVSSPP 753

Query: 124 APQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183
 A P+ S ++SP PAP S A ++K+ + + P PP + PP +P
 Sbjct: 754 A-LAPVSSPPSVKSSPPAPLSSPPPAQVKS---SPPPVQVSSP--PPAPKSSPPLAP 806

Query: 184 ASTASFIFSKGSRKLQLERP-VSPETQADLQRNLVAELRSISEQRPPQAPKKSPPKAPPPV 242
 S+ + L P ++P++ +V+ + + PP AP SP P
 Sbjct: 807 VSSPPQVEKTSPPAPLSSPPLAPKSSPP--HVVVSSPPPVKSSPPPPAPVSSPPLTPKP 864

Query: 243 ARKPS-VGVPP----PASPSYPR-----AEPLTAPP 268
 A P+ V PP P++P P +EP ++PP
 Sbjct: 865 ASPPAHVSSPPEVVKPSTPPAPTTPVISPPSEPKSSPP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPEL-VSSP--AASSSSATALQIQPPG--SPDPPPPAP--- 56
 PPP A S P P S P + VSSP A SS A PP PPPAP
 Sbjct: 768 PPP--APLSSPPAPQVKSSPPPVQVSSPPAPKSSPPLAPVSSPPQVEKTSPPAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPEGGAP 116
 P AP SS P V P PV S PP V +P +TP V +P
 Sbjct: 826 PPLAPKSSPPHHVVSSPP--PVVKSS---PPAPVSSPPLTPKPASPPA--HVSSPPEVV 878

Query: 117 TPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQP--- 169
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPATTVISPPSEPKSSPPPTFVSLPPPVIKSSPPAPMVSSPPMTPKSSPPPVVV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIFSKGRKLQLERPVS PETQADLQRLNVAELRSISEQRP 228
+ PP + PP + P S + + P PE ++ V + + P

Sbjct: 938 SSSPPTVKSSPPAPVSSPPATPKSSPPAPVNL P-PPEVKSSPPPTPVSSPPAPKSSP 996

Query: 229 PQAPKKS PKAPPVARKPS---VGVPASPSPSYPRAEPLTAPP 268
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPPEVKSSPPAPVSSPPPVKSSPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGPSGSELVSSPAASSSSATALQIQPP---GSPDPP---PA 55
D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKL---PQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVLRLSVGA 111
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPPVSVSSPPPVKSSPPAPVGSPPPEKSSPPAPVASPPPVKSSPP 584

Query: 112 PG--GAPTALGPSAPQKPLRRA---LSGRASVPAPSSGLHAAVRLKACSLAASEGLSS 166
P + P P + P P + + P P S A V + + +

Sbjct: 585 PTLVASPPPVKSSPPAPVASPPPVKSSPPPTPVASPPPPAPVASSPPPMKSSPPPTP 644

Query: 167 AQPNGPPEAEPRPPQSPASTASFIFSKGRKLQLERPVS PETQADLQRLNVAELRSISEQ 226
PPE P PP PA + + ++ PE L+ +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-PTTLIPSPPPQEK 702

Query: 227 RPPQAPKKS PKAPP-PVARKPSVGVPASPSPSYPRAEPLTAPP 268
PP P K P + P P K V PP S P P+++PP

Sbjct: 703 TTPSTPSKPPSSPEKSPPKPEVSSPPQTPKSSPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAEFFSVASPEPAGP---SGSELVSSPAASSSSATALQIQPPGSP--DPPAP-- 56
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPPAPVSSPPLTPKPAASPAPVSSPPEVVK-PSTPPAPTIV--ISPPSEPKSSPPPTPV 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVLRLSVGAPGGA 115
P P SS P + P P PP V + P P++ V + P

Sbjct: 907 LPPPVIKSSPPAPMVSSPPMTPKS-----SPPPVVSSP--PPTVKSSPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 PKSSPPAPVNLPPPEV---KSSPPPTPVSSPPAPK----SSPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIFSKGRKLQLERPVS PETQADLQRLNVAELRSISEQRPQAPKKS 235
+ PP + P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPAPVSSPPPVKSSPPAPVSSP-PPPVKSSPPAPVSSPPPVKSSPPAPISS 1068

Query: 236 PKAPPVARKPS---VGVPASPSPSYPRAEPLTAPP 268
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P---PPPVKSSPPAPVSSPPPVKSSPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPPGSPDPPAP---PAPAPASSAPGHV 69
P P G P SP + PAAS+ S T + P P+P P P P P P +P

Sbjct: 410 PTPGGGPSSP-VPGKPAASAPMPSPTTPDVSPPELPEPSVPAPAPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLSLQMVLRLSVGAPGGAPTALGPSAPQK 128
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVSSPPPVKSSPPAPVG---SPP--PPEKSPPPAPVASPP 577

Query: 189 FIFSKGRKLQLERPVS---SPETQADLQRLNVAELRS-----ISEQRPPQA-----PK 233
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSSPPPTLVASPPPVKSSPPAPVA-SPPPVKSSPPPTPVASPPPPAPVASSPPPM 636

Query: 234 KSPKAPPVARKP---SVGVPASPSPSYPRAEPLTAPPTN 270
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPPPTPVSSPPPEKSSPPPPAKSTPPPEEYPTPTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFSSVASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSPDPPAPPAPAPASS 64
 PP S S + P +P + P SS A+ PP +P +PP P SS
 Sbjct: 883 PPAPTIVISPPSEPKSSPPPTFVSLPPPVIKSSPPAMPVSSPPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVVGAPG--GAPTALGP 122
 P V P PV PP +P P L ++S P +P PA
 Sbjct: 940 PPPTVKSSPPAPVS-----SPPATPKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP 180
 S P P+ ++ P PAP S V+ S +SS P PP + PP
 Sbjct: 995 SPPAPMSSPPPEVKSSPPAPVSSPPPPVK----SPPPPAPVSS--P--PPPVKSPPP 1046

Query: 181 QSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKSPKAPP 240
 +P S+ + P P ++ V+ + PP AP SP PP
 Sbjct: 1047 PAPVSSPPPPVKSPPPAPISSP-PPPVKSPPPAPVSSPPPPVKSPPPAPVSSP--PP 1103

Query: 241 PVARKPS---VGVPPPAS---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283
 P+ P V PPA PS P P+++P P + ++ L
 Sbjct: 1104 PKSPPPPPAPVSSPPAPVKKPSLPPAPVSSPPPVTPAPPKKEEQL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFSSVASPEPAGPSGSPELVSSPAASSSSATALQIQPP----GSPDPP---PA 55
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +
 Sbjct: 469 DVVPPTPP---VPGKSPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTTS 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVVGAPGGA 115
 PPAP + S P V+ + PV PP VG+P P V +P
 Sbjct: 525 PPAPIGSPSPPPFVSVVSPPPVKSP----PPAPVGSPP--PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEA 175
 P P P P ++ P PAP + V+ S ++S P P +
 Sbjct: 576 PPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVK----SPPPTPVASPPPPAPVAS 631

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
 P P +SP K P P S+ PP+
 Sbjct: 632 SPPPMKSPPPPTPVSSPPPEKSP--PPPPAKSTPPPEEYPTPTTSVKSSPPPEKSLFP 689

Query: 236 PK---APPPVARK--PSVGVPPPASPSYPRA--EPLTAPP 268
 P +PPP + PS PP+SP P EP+++PP
 Sbjct: 690 PTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPPEEAFSSVASPE-PAGPSGSPELVSSPAASSSSATALQIQPPGSPDPP-PAPPAP 59
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P
 Sbjct: 496 ASTPPP--SLVKLSPPQAPVGSPPPPVKTTSPAPIGSPSPPPVSVSSPPPPVKSPPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVVGAPGAPTAP 119
 AP S P P PV PP + P + S V+ AP +P P
 Sbjct: 554 APVGSPPPEKSPPPAPVASPP--PPVKSPPPPTLVASPPPPVKSPPPAPVASPPPP 610

Query: 120 LGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPR 178
 + P P+ + P PAP + ++ +S P PP A+
 Sbjct: 611 VKSPPPPTPVA-----SPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKST 664

Query: 179 PP--QSPASTASFIFSKGSRKLQLERPV---SPETQADLQRNLVAELRSISEQRPPQAPK 233
 PP + P S S S K L P SP Q S ++P +P
 Sbjct: 665 PPPEEYPTPTTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKP--SPP 721

Query: 234 KSPKAPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 K P + PP K S PPA S P P+++PP
 Sbjct: 722 KEPVSSPPQTPKSS---PPAPVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGPSGSP-LSVSSPAASSSS---ATALQIQPPGSPDPPP-- 54
 PPPE++ VASP P S P LV+SP S A PP PPP
 Sbjct: 560 PPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVKSPPPPTP 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVGC---SKGGGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP +S+P + P PV K PP P ++S
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTTSVK 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSA 167
 P + P P L PS P P + + ++P PSS + + S SS
 Sbjct: 680 SPPPEKSLPPPTLIPSP--PQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSP 736

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQR 227
 P P SP + A + S S K P + P + + +
 Sbjet: 737 PPAPVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPPVQVSS 793

Query: 228 PPQAPKKSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 PP APK SP P+A P V PP + P PL++PP
 Sbjet: 794 PPPAPKSSP----PLA--P-VSSPPQVEKTSPPAPLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09
 Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFSSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQPPGSPDPPP-APPAPA 60
 PPP + + + P P G PS P +VS P S P GSP PP +PP PA
 Sbjet: 517 PPPVK--TTSPPAPIGSPSPPPVSVVSPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP---LVTPSLLQMVRLRSVGAPGG 114
 P +S P V P V PP V +P + +P V AP
 Sbjet: 571 PVASPPPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASPVPAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171
 + P + P P+ SP P P S+ S+ +S + P
 Sbjet: 631 SSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTPSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQA 231
 PP P PP T SK P SPE + + V+ + PP A
 Sbjet: 689 PPTLIPSPPPQEKTPPSTPSKP-----PSSPEKSP-PKEPVSSPPPTPKSSPPPA 739

Query: 232 PKKSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268
 P SP P PV+ P++ PP+ S P PL++PP
 Sbjet: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPAPLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08
 Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFSSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQPPGSPDPPPAPPAPA 60
 A P P S PEP PS P P + S A PP P P +PPA +
 Sbjet: 427 ASAPMPSHTPPDVSPEPLPEPSVPAPAPMMPPTPHSPADDVVPPTPPVPGKSPATS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT-- 118
 P+ A P V S PP+ VG+P P V+ S AP G+P+P
 Sbjet: 487 PSPVQVPPAASTPPPSLVKLS----PPQAPVGSPP--PPP----VKTTSPAPIGSPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
 + P P K P A G SP P S A S + + PP
 Sbjet: 537 PVSVSSPPPPVKSPPPAPVG--SPPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPP 594

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKK 234
 + PP +P ++ + P P A + + PP P+K
 Sbjet: 595 VKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPPTNGLP 273
 SP PPP P PP P+ P + + PP LP
 Sbjet: 654 SPPPPPPAKSTP---PPEEYPTPTPSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08
 Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFSSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQPPGSP--DPPPAP---PAP 59
 PPP V+SP P P SP P SS ++ PP +P PP P P P
 Sbjet: 916 PPPA---MVSSP-PMTKSSPP---PVVSSPPPTVKSSPPAPVSSPPATPKSSPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT 119
 AP + P V P PV S P AP+ +P + V+ AP +P P
 Sbjet: 967 APVNLPPPEVKSSPPPTVVS-SPPAPKSSPPAPMSSPPPPPE-VKSPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEG---LSSAQPNGPPEA 175
 + P P+ ++ P PAP S V+ S + S P P +
 Sbjet: 1025 VKSPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKK 235
 P P +SP A S ++ P P A + A ++ S PP AP S
 Sbjet: 1085 PPPPVKSPPPAPV---SSPPPIKSPPPP---APVSSPPAPVKPPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYPRAEPLTAPP 268
 P P +K +PPPA S P + PP
 Sbjet: 1136 PPPVVTAPPKKEEQLPPPAESQPPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06
 Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFFPPPEEAFSSVASPEP-AGPSGSELVSSPAASSSSATA-LQIQPPGSP--DPPP---A 55
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +
 Sbjct: 970 NLPPPEVK--SSPPPTPVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSSPPPPVKS 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115
 PP PAP SS P V P PV PP + P S V+ AP +
 Sbjct: 1028 PPPPAPVSSPPPPVKSPPPPAPVSSPP--PPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 116 PTPALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
 P P + P P+ ++ P PAP S A +K SL +SS P PP
 Sbjct: 1085 PPPPVKSPPPPAPVSSPPPIKSPPPAPVSSPPPPAP-VKPPSLPPAPVSS--P--PPV 1139

Query: 175 AEPRPPQ 181
 P PP+
 Sbjct: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPPEEAFSSVASPEPAGP-SGSELVSSP---AASSSSATALQIQPPGSP--DPPP 54
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1001 MSSPPPE-----VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPP 1055

Query: 55 ---APPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP SS P V P PV PP V +P P +
 Sbjct: 1056 PVKSPPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125
 V +P AP P+L P AP
 Sbjct: 1114 VSSPPPPAPVKPPSLPPPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFSS---VASPEPAGP-SGSELVSSP---AASSSSATALQIQPPGSP--DPPP 54
 PPP S V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1060 PPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPPIKSPPPAPVSSPPPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108
 AP P PAP SS P V P K+ + PP E P +L +
 Sbjct: 1120 APVKPPSLPPPPAPVSSPPPVTPAPPKKE---EQSLPPAESQPPPSFNDIILPPIMANK 1176

Query: 109 VGAP 112
 +P
 Sbjct: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQPNG 171
 G PTP GP + P + A S +P+P+P + L S + A + P+
 Sbjct: 408 GYPTPGGGPPSSPVGKPAAS---APMPSHTPPDVSPLEPLPEPSVPAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFIFSKGRKLQLERPVSPETQ---ADLQRNLVAELRSISEQR 227
 PP + PP P S + S ++Q +P + Q + + +
 Sbjct: 465 PPADDDYVPFTPPVPGKSPATSPSPQPPAASTPPFSLVKLSPPQAFVGSPPPPVKTTS 524

Query: 228 PPQAPKSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 PP AP SP PPPV SV PPP S P P+ +PP
 Sbjct: 525 PP-APIGSPSPPPV----SVVSPPPPVKSPPPPAPVGSPP 560

Pedant information for DKFZphmcf1_lc23, frame 1

Report for DKFZphmcf1_lc23.1

[LENGTH]	311
[MW]	31534.58
[pI]	9.48
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 38.59 %

SEQ	MADFPPPEEAFSSVASPEPAGPSGSELVSSPAASSSSATALQIQPPGSPDPPPPAPPAPA
SEGxx
PRD	cc
SEQ	PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTAL
SEG	xxxxxx.....xx


```
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccc
SEQ      GPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPFEAEPRPP
SEG      xxxxx.....xxxxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      QSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKKS PKAPP
SEG      xxxxx.....xxxxxxxxxxxxxxxx
PRD      cccccceeeccccchhhhhccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      PVARKPSVGVP PPASPSYPRAEPLTAPPTNGLPHTQDRTKRELAENGGVLQLVGPEEKMG
SEG      xxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhcccccecccccccccc

SEQ      LPGSDSQKELA
SEG      .....
PRD      ccccccccccc
```

(No Prosite data available for DKFZphmcf1_lc23.1)

(No Pfam data available for DKFZphmcf1_lc23.1)

DKFZphmcf1_1e15

group: transmembrane protein

DKFZphmcf1_1e15 encodes a novel 454 amino acid protein with similarity to *C. elegans* proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER
membrane regions: 9

complete cDNA, complete cds, EST hits
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp

Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGACGCGC  CCGGGCTGAG  CGACAGCAAG  TGCAGCGGGC  TCCTACCCCG
51  GGTGAGGGGT  GGCCTCCGCG  TGGGATCGTG  CCCTCTTCAG  CCCGCTCCTG
101  TCCCGACAT  CACGTGTATT  CCGCACGTCC  CCTCCGCGCT  GTGTGTCTAC
151  TGAGACGGGG  AGCGGTGACA  GGGCCCGGGT  CCCTTCTCAG  TGGTGTCTG
201  TGCTTCAGGG  CAAGCTCCCC  GTCTCCGGGC  GCACCTCCCT  CGCCTGTGTT
251  CGGTCCATCC  TCCTTTCTCC  AGCCTCCTCC  CCTCGCAGGT  GGGATCGTCG
301  GTGGGACCGG  AGCGCGGGCG  GGCGCGGGCC  CCCGGGACCA  TGGCCGGGTC
351  CGACACCGCG  CCCTTCCTCA  GCCAGGCGGA  TGACCCGGAC  GACGGGCCAG
401  TGCCTGGCAC  CCCGGGGTTG  CCAGGGTCCA  CGGGGAACCC  GAAGTCCGAG
451  GAGCCCGAGG  TCCCGGACCA  GGAGGGGCTG  CAGCGCATCA  CCGGCCTGTC
501  TCCCGGCGGT  TCGGCTCTCA  TAGTGGCGGT  GCTGTGCTAC  ATCAATCTCC
551  TGAACACTAC  GGACCGCTTC  ACCGTGGCTG  TGTTCATCTC  CAGTTACATG
601  GTGTGGCAC  CTGTGTTTGG  CTACCTGGGT  GACAGGTACA  ATCGGAAGTA
651  TCTCATGTGC  GGGGGCATTG  CCTTCTGGTC  CCTGGTGACA  CTGGGGTCAT
701  CCTTCATCCC  CGGAGAGCAT  TTCTGGCTGC  TCCTCCTGAC  CCGGGGCCCTG
751  GTGGGGGTGC  GGGAGGCCAG  TTATTCACCC  ATCGCGCCCA  CTCTCATTGC
801  CGACCTCTTT  GTGGCCGACC  AGCGGAGCCG  GATGCTCAGC  ATCTTCTACT
851  TTGCCATTCC  GGTGGGCAGT  GGTCTGGGCT  ACATTGCAGG  CTCCAAAGTG
901  AAGGATATGG  CTGGAGACTG  GCACTGGGCT  CTGAGGGTGA  CACCGGGTCT
951  AGGAGTGGTG  GCCGTTCTGC  TGCTGTTCT  GGTAGTGCGG  GAGCCGCCAA
1001  GGGGAGCCGT  GGAGCGCCAC  TCAGATTTGC  CACCCCTGAA  CCCCACCTCG
1051  TGGTGGGCG  ATCTGAGGGC  TCTGGCAAGA  AATCTCATCT  TTGGACTCAT
1101  CACCTGCCTG  ACCGGAGTCC  TGGGTGTGGG  CCTGGGTGTG  GAGATCAGCC
1151  GCCGGCTCCG  CCACTCCAAC  CCCCGGGCTG  ATCCCTGGT  CTGTGCCACT
1201  GGCCTCCTGG  GCTCTGCACC  CTTCCTCTTC  CTGTCCCTTG  CCTGCGCCCG
1251  TGGTAGCATC  GTGGCCACTT  ATATTTTCAT  CTTTCATTGA  GAGACCTCTC
1301  TGTCCATGAA  CTGGGCCATC  GTGGCCGACA  TTCTGCTGTA  CGTGGTGATC
1351  CCTACCCGAC  GCTCCACCGC  CGAGGCCTTC  CAGATCGTGC  TGTCCACCT
1401  GCTGGGTGAT  GCTGGGAGCC  CCTACCTCAT  TGGCCTGATC  TCTGACCGCC
1451  TGCGCGGAA  CTGGCCCCC  TCCTTCTTGT  CCGAGTCCG  GGTCTGTCAG
1501  TTCTCGCTCA  TGCTCTGCGC  GTTGTGTGG  GCACTGGGCG  GCGCAGCCTT
1551  CCTGGGCACC  GCCATCTTCA  TTGAGGCCGA  CCGCCGGCGG  GCACAGCTGC
1601  ACGTGCAGGG  CCTGCTGCAC  GAAGCAGGGT  CCACAGACGA  CCGGATTGTG
1651  GTGCCCCAGC  GGGGCCGCTC  CACCCGCGTG  CCGTGGCCA  GTGTGCTCAT
1701  CTGAGAGGCT  GCCGCTCACC  TACCTGCACA  TCTGCCACAG  CTGGCCCTTG
1751  GCCACCCCA  CGAAGGGCCT  GGGCCTAACC  CCTTGGCCTG  GCCCAGCTTC
1801  CAGAGGGACC  CTGGGCCGTG  TGCCAGCTCC  CAGACACTAC  ATGGGTAGCT
1851  CAGGGGAGGA  GGTGGGGGTC  CAGGAGGGGG  ATCCCTCTCC  ACAGGGGCAG
1901  CCCCAGGGC  TCGGTGCTAT  TTGTAACGGA  ATAAAATTG  TAGCCAGAAA
1951  AAAAAAA

```

BLAST Results

Entry E12646 from database EMBL:
cDNA encoding cell growth inhibiting factor.
Score = 3046, P = 2.2e-131, identities = 640/659

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454
 Category: similarity to known protein

```

1 MAGSDTAPFL SQADDPDDGP VPGTPGLPGS TGNPKSEEP VPDQEGLRQI
51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISYVMVLA PVFGYLGDRY
101 NRKYLKCGGI AFWSLVTLGS SFIPGEHFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRMSIFY FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVAVL LFLVVRPEP RGAVERHSDL PPLNPTSWSA DLRLARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRA DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCFAVGA LG
401 GAAFLGTAIF IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcfl_1el5, frame 1

TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,
 N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid
 C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,
 N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11_18 gene: "F6H11.180"; product: "predicted protein";
 Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII
 project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER) ., N
 = 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9
 Length = 488

HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISSYVMVLA PVFGYLGDRYNRKYLKCGGIAFWSLVT 117
 + ++ V Y N + + + VF+ S+MV +PV GYLGR+NRK++M G+ W
 Sbjct: 29 AGVLTQVQTYYNISDSLGLLIQTVFLISFMVFPVCGYLGDRFNKRWIMIGVGIWLGAV 88

Query: 118 LGSSFIPGEHFWL LLLTRGLVGVGEASYSTIAPT LIADLFVADQRSRMSIFYFAIPVGS 177
 LGSSF+P HFWL L+ R VG+GEASYS +AP+LI+D+F +RS + IFYFAIPVGS
 Sbjct: 89 LGSSFVPANHFWLFLVLRSFVGIGEASYSNVAPSLISDMFNGQKRSTVEMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLL LFLVVRPEPRGAVER----HSDLPPL 233
 GLG+I GS V + G W W +RV+ G++ ++ L L EP RGA ++ D+
 Sbjct: 149 GLGFIVGSNVATLTGHQWQIRVSAIAGLIVMIALVLTFTYEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWWADLRALARNLIFGLITCLTG 259
 T++ DL L + L+ C G
 Sbjct: 209 TNTTYLEDLVILLKTPT--LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGV LGVGLGVEISRRL-----RHSNPRADPLVCATGLLGSAPFLFLSL 300
 L FG IT G++GV G +S+ L R RA PLV G L +APFL + +
 Sbjct: 277 LYFGAITTAGGLIGVIFGSMKWL VAGWGPFRRLQTDRAQPLVAGGGALLAAPFLIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

Score = 70 (10.5 bits), Expect = 5.9e-24, Sum P(2) = 5.9e-24
Identities = 25/89 (28%), Positives = 41/89 (46%)

Report for DKFZphmcf1_1e15.1

[BLOCKS]	BL01022D	
[PROSITE]	MYRISTYL	11
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	3
[PROSITE]	PROKAR_LIPOPROTEIN	1
[PROSITE]	GLYCOSAMINOGLYCAN	1
[PROSITE]	PKC_PHOSPHO_SITE	4
[KW]	TRANSMEMBRANE	8
[KW]	LOW_COMPLEXITY	15.42 %

SEQ HVQGLLHEAGSTDDRI VVPQGRGRSTRVPVASVLI


```

SEG .....
PRD hhhhhhhhhcccccccccccccccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

Prosites for DKFZphmcf1_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1e15.1)

DKFZphmcf1_lg13

group: mammary carcinoma derived

DKFZphmcf1_lg13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKFZ

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```
1 GAAACCTGAT CTCATAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51 GACCCTATTT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACTACA GATTGTCTC CAGGGGGTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGT CTCAGTCGCT
251 CTACAACTAT GAATGAGAGA GCCTTATTGT CATCGTATTT AGTTGCATAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA GCGGCTGAAA AAATTATCCT
351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATTT GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCCACT CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
651 TTTATTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAAACT
701 GGAACATTG TAAAGGAATT TCAAGTGATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACAATGC
801 TGTTTGAAT CACTGTTTGA TTCATCGAGA AGCTTTGGTA TCCAAGAAA
851 TTTCAACAAG TCTGATGGAT GTATTGAAAA ATGCAAGTAA AACTGTTAAT
901 TTTATTAAG GAAGCTCACT GAATAGCCGA CTTCTGAAA TATTTTGTTT
951 AGAGATTGGA GTGAACCACA CCCACTTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAC CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTTGGCAA ATATTTTGA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTGAT ATTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAAAACAA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGCTCAAA CTTTGAATTA
1401 TTAATTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAAT CAATAATTGA GTTAACTTG
1501 GAGCCTGAAG AAGAGAATGA ATTATTGCAG CTCAGTTCAT CATTCACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACTA CATATTTGTG TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAAAG AAGAGAAATA GGCTCAATAG TGCACCATG ATGCGGGTAG
1751 CATTATCTTC ATGTGTTTCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCCATCAC ATTAATACA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAAC CCCATCTCTA
1951 CTAATAATAG AACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCCTGAGGC AGGAGAATCT CTTAAACCAG GAAGGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGCTGA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAAATTGT ATTTGTAATT TTAAGGGAT
2151 TTTGCAGTAT GTTGAGTTA AACGTTAATA AAATTATATT TGAATTAGG
2201 AAAAAA
```

BLAST Results

Entry AC005020 from database EMBL:
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.
Score = 9110, P = 0.0e+00, identities = 1822/1822

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573
 Category: similarity to unknown protein

```

1 MTPESRDTTD LSPGGTQEME GIVIVKVEEE DEEDHFOKER NKVESSPQVL
51 SRSTTMNERA LLSSYLVAIR VAKEKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKHCKGIS SDGTANMTGK HSRLTEKLE ATHNNAVWNH CFHREALVS
251 KEISPSLMDV LKNAVKTVMF IKGSSLNSRL LEIFCSEIGV NHTHLLFHTE
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSOI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFQK TLLWQARLK SNRPSYMFEP
401 TLLQHTEENI INEDCLKEIK LEILLHLTSL SQTENYFFPE EKFECLKENI
451 WMKDPFAFQN PESIIELNLE PEEENELLQL SSSFTLKNNY KILSLSAFWI
501 KIKDDFFLLS RKSILLLLPF TTYLCELGF SILTRLKTKK RNRLNSAPDM
551 RVALSSCVDP WKELMNRQAH PSH

```

BLASTP hits

Entry AC004877_3 from database TREMBLNEW:
 gene: "WUGSC:H_DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens
 PAC clone DJ0751H13 from 7q35-qter, complete sequence.
 Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211_1 from database TREMBL:
 product: "Hermes transposase"; Musca domestica Hermes transposase
 gene, complete cds.
 Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

Alert BLASTP hits for DKFzphmcf1_1g13, frame 1

TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
 sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P
 = 1.1e-23

>TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
 sapiens mRNA for KIAA0766 protein, complete cds.
 Length = 607

HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23
 Identities = 120/485 (24%), Positives = 229/485 (47%)

```

Query:      89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
             CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+
Sbjct:     124 CMEVLLREVLPEH-VSVLQGVDLSPDITRQILSIDRNLRLNQFNRRARDFKAYSLALDDQ 182

Query:     148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
             +A LLV++R V + + EDLL +NL H + G + LE+ L L+ +
Sbjct:     183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query:     206 CKGISSDGTANMTGKHSRLTEKLEATHNNAVWN--HC--FIHREALVSKEISPSLMDVL 261
             G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct:     241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPNCWNVIHYSGLFLHLELLSSYDQVDVN--QII 298

Query:     262 KNAVKTVMFIKSSSLNSRLLEIFCSEIGVNHHTHLLFHTEVR-WLSQGVKLSRVYELRNEI 320
             + + IK + + +E H + + WL +GK L ++ LR E+
Sbjct:     299 NTISEWIVLIKTRGVRRPEFQTLTESESEHGERVNGRCLNNWLRGKTKLKFSLRKEM 358

Query:     321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFGILNELSLKMQGKNNDIFQYLEHILGFQK 380
             FLV + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:     359 EAFVLSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417

```



```

Query:      89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRCTIAKHEAMLITRQSGIDFAIQDES 147
            CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+
Sbjct:     124 CMEVLLREVLPEH-VSVLQGVLDSPDITRQIRLSIORNLNRQLFNRRDFKAYSLALDDQ 182

Query:     148 TDIASCPTLLVYRVYVQD-DFVEDLLCLLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
            +A LLV++ R V + + EDLL +NL H + G + Le+ L L+ +
Sbjct:     183 AFVAYENYLLVIRFGVPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query:     206 CKGISSDGTANMTGKHSRLTEKLEATHNNNAVWNHCFIHREALVSKEISPSLMDV-LKNA 264
            G+++ T M G++S L + E + WN IH + E+ S DV +
Sbjct:     241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPNCWN--VIHYSGLFHLELLSSY-DVDVNQI 297

Query:     265 VKTVN----FIKGSSLSNRLLIEFCSEIGVNHTHLLFHTEV-RWLSQGKVLRSRYELRNE 319
            + T++ IK + + +E H + + WL +GK L + + LR E
Sbjct:     298 INTISEWIVLIKTRGVRPEFQTLTLTESESHGERVNGRCLNWLRRGKTLKLIFSRLRE 357

Query:     320 IYIFLVEKQSHLANIFEDDIWVTKLAYLSDFIGILNELSLKMQGKNNDIFQYLEHILGFQ 379
            + FLV + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:     358 MEAFVLSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFE 416

Query:     380 KTLWLQARLKS NRPSYMFPTLLQHIEENIINEDCLKEIKL----EILLHLTSLSQTFN 435
            L L+Q ++ + FP L ++E + + ++ K+ + + L + F
Sbjct:     417 VKLNLFRHIEEKNLTD--FPALREVDE--LKQNKDEKIFDPDRYQMVICRLQKFE 472

Query:     436 YYPFEKFESLKENIWM-KDPFAFQNPESIIELNLEPEEENELLQSSSFTLNKYYKILS 494
            +F + + F +K+++ + +PF F+ + I + +E L +L ++ L N Y+I
Sbjct:     473 RHFKDLRF--IKKDLELFSNPFNFKEPYAPISVRVE-----LTKLQANTNLWNEYRIKD 524

Query:     495 LSAFWIKIK-DDFPLLSRKSILLLPFTTYYLCELGFSLTRLTKTKKRNRLNSA---PDM 550
            L F+ + + +P++ + + F + +CE FS LTR + L .
Sbjct:     525 LGQFYAGLSAESYPIIKGVACKVASLFDNSQICEKAFSYLTRNQHTLSQPLTDEHLQALF 584

Query:     551 RVALSSCPVDWKELMNRQAHPH 573
            RVA + P W +L+ R+ + S+
Sbjct:     585 RVATTEMEPGWDDLVR-RERNESN 606

```

Report for DKFZphmcf1_lg13.1

```

SEQ      MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDDEEDHFQKERNKVESSPQVLSRSTTMNERA
SEG      .....xxxxxxx.....
PRD      cccccccccccccccccceeeeeeeccccchhhhhhhhhhhccccceeeccchhhh
SEQ      LLSSYLVAIRVAKEKMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRC

```



```

SEQ      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhccccceeeeeecccchhhhhh

SEQ      TI AKHLEAMLITRLQSGIDFAIQ LDESTDIASCPTLLVYVRVWQDDFVEDLLCCLNLS
SEG      .
PRD      hhhhhhhhhhhhhhhhhheeeccccccccccccccceeeeeecchhhhhhhhhhcce

SEQ      HITGLDLFTELENCLLGQYKLNWKHKCGISSDGTANMTGKHSRLTEKLEATHNNAVWNH
SEG      .
PRD      eeeehhhhhhhhhhhhhhhccccccccccccccceeecccchhhhhhhhhhhcccccee

SEQ      CFIHREALVSKEISPSLMDVLKNAVKTVNFIKGSSLNSRLEIFCSEIGVNHTHLFFTE
SEG      .
PRD      hhhhhhhhhhhhhccccchhhhhhhhhhhheeeccccchhhhhhhhhhhccccchhhhhhh

SEQ      VRWLSQGKVL SRVYELRNEIYIFLVEKQSHLANI FEDIWVTKLAYSDFIGILNELSLK
SEG      .
PRD      ccccccccchhhhhhhhhhhhhhhhhhhhhhhchhhhhccccceehhhhhhhhhhhhhhhhh

SEQ      MQGKNNDIFQYLEHILGFQKTLLLQWARLKSNRPSYMFPTLLQHIEENIINEDCLKEIK
SEG      .
PRD      hhcccccccchhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhcchhhhhh

SEQ      LEILLHLTSLSQTFNYFYFPEEKFESLKENIWMKDPAFQNPESIIELNLEPEEENELLQL
SEG      xxxxxx.....xxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhccccccccceeecccchhhhhhhhh

SEQ      SSSFTLKNYYKILSLSAFWIKIKDDFP LLSRKSILLLPFTTTYLC ELGFSILTRLTKTK
SEG      xxx.....xxxxxxxxxxxxx
PRD      hhccccchhhhhhhhhhhhhccccccccccchhhhhhhhhccccceehhhhhhhhhhhhhhh

SEQ      RNRLNSAPDMRVALSSCPVDWKELMNRQAHP SH
SEG      .
PRD      hccccccccceeeccccccchhhhhhhcccccc

```

Prosites for DKFZphmcf1_1g13.1

PS000001	216->220	ASN_GLYCOSYLATION	PDOC00001
PS000001	291->295	ASN_GLYCOSYLATION	PDOC00001
PS000005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS000005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS000005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS000005	391->394	PKC_PHOSPHO_SITE	PDOC00005
PS000005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS000005	485->488	PKC_PHOSPHO_SITE	PDOC00005
PS000005	510->513	PKC_PHOSPHO_SITE	PDOC00005
PS000005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS000006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS000006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS000006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS000006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS000006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS000006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC00006
PS000006	445->449	CK2_PHOSPHO_SITE	PDOC00006
PS000006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS000006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS000007	364->372	TYR_PHOSPHO_SITE	PDOC00007
PS000008	137->143	MYRISTYL	PDOC00008
PS000008	273->279	MYRISTYL	PDOC00008
PS000008	289->295	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphmcf1_1g13.1)

DKFZphtes3_14g5

group: testes derived

DKFZphtes3_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and nuclear localization signals of lyar.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```

1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTCGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTACCTTA AAAATGGTAT
151 TTTTACATCG CAATGCATGT GGTGAATCAG TGAAGAAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGCAG AAACGTGTAA TGCCTTTCTT GCATTGACTG
251 CGGTAAAGAT TTCTGGGGCG ATGACTATAA AAACCACGTG AAATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAGGGCT ATGAAGGTAA AACCCACAAA
351 GGCACATCA AACAGCAGGC GTGGATTGAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCCA AAGTGAGAGA ACTTTTAGAG CAAATTAGTG
451 CTTTGTGACA CGTTCCCAGG AAAAAGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTCTGTA GCTTCCAACA GCGAACCACT CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAAGG AAAAGAGAAA
751 AGAAAGAACT AAAGTTAGAA AACCACCAGG AAAACTCAAG GAATCAGAAG
801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTTGAGG CTGGTGGGGA
851 GGAAGTCCCT GAGGCCAATG GCTCTGCAGG GAAGAGGAGC AAGAAGAAGA
901 AGCAGCGCAA GGACAGCGCC AGTGAGGAAG AGGCACGCGT GGGCGCAGGG
951 AAGAGGAAGC GGAGGCACAT GGAAGTTGAA ACAGATTCTA AGAAGAAAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGGCGG AGAACCAGAA GACGATGAGG
1051 CTCCTGCAAA AGGTAAATTC AACTGGAAGG GAACTATTAA AGCAATTCTG
1101 AAACAGGCCC CAGACAATGA AATAACCATC AAAAAGCTAA GGAAAAAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCACAGA TCCGAAGAGG
1201 AACTCCTGGT CATCTTAAAC AAGAAAATCA GCAAGAACCC TACCTTTAAG
1251 TTATTAAAGG ACAAAGTCAA GCTTGTGAAA TGAACATTG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTTCTTCTT TCACTGCTGT TTATAAAATG
1351 TGTAATGAAT TCTAACAAC CAAATTTTGC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAAATAT
1451 ATTCTGGTCC AAACCTCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1501 AAA

```

BLAST Results

No BLAST result

Medline entries

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379
 Category: strong similarity to known protein
 Classification: Cell division
 Prosite motifs: ATP_GTP_A (60-68)

```

1 MVFFTCNACG ESKKKIQVEK HVSVCNCEC LSCIDCGKDF WGGDYKNHVK
51 CISEDQKYGG KGYEGKTHKG DIKQQAQWIK ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNWMMN SLKVHNESIL DQVWNIFSEA SNSEPNKEQ
151 QDRPLHPVAN PHAEISTKVP ASKVKDAVEQ QGEVKKNKRE RKEERQKKRK
201 REKKELKLEN HQENSRNQKP KKKKKQGEAD LEAGGEEVPE ANGSAKGRSK
251 KKKQKDSAS EEEARVGAGK RKKRHSEVET DSKKKKMKLP EHPEGGEPEP
301 DEAPAKGKFN WKGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHHRS
351 EEELLVIFNK KISKNPTFKL LKDKVKLVK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58 CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III, N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058 18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse
 Length = 388

HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144
 Identities = 275/388 (70%), Positives = 317/388 (81%)

```

Query:      1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYGG 60
            MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGGDYK+HVKCISE QKYGG
Sbjct:      1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGGDYKSHVKCISEGQKYGG 60

Query:      61 KGYEGKTHKGDIKQQAQWIKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNWMMN 120
            KGYE KTHKGD KQQAQWIKI+ELIK+PNVSPKVRELL+QISAFDNVP KAKFQNWMMN
Sbjct:      61 KGYEAKTHKGDAKQQAQWIKINELIKPNVSPKVRELLQQISAFDNVPIKKAKFQNWMMN 120

Query:      121 SLKVHNESILDQVWNIFSEASNSEPNKEQDQRPLHPVANPHAEIS-TKVPASKVKDAVE 179
            SLKVH++S+L+QVW+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct:      121 SLKVHSDSVLEQVWDIFSEASSSE---QQQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query:      180 QQGEVKKNKRRERKEERQKKRKREKKELKLENHQENSRNQPKKKKKQGEADLEAGGEEVP 239
            +Q E KKNKRERKEERQK RK+EKELKLENHQEN R QPKKKRK QEA EA GE+
Sbjct:      177 EQTEAKKNKRERKEERQKNRKKKKELKLENHQENLRGQPKKKRKNQEAHEAAGEDGA 236

Query:      240 EANG-----SAGKRSKKKKQKQKDSASEEEA----RVGAGKRRR-RHSEVETDSKKKKM 287
            + +G G+ S++ R E+ A + AGKRRR +HS E+ KKKKM
Sbjct:      237 DSGGPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRRRPHKSGAESGYKKKKM 296

Query:      288 KLPEHPEGGEPEDEAPAKGKFNWKGTIKAILKQAPDNEITIKLKKVLAQYTVTDEH 347
            KLPE PE GE +D EAP+KGKFNWKGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct:      297 KLPEQPEEGEAKDHEAPSGKFNWKGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query:      348 HRSEELLVIFNKKISKNPTFKLLKDKVKLVK 379
            EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct:      357 SHHEELLAIIFNRKISRNPFTFKVLKDRVKLLK 388

```

Pedant information for DKFZphtes3_14g5, frame 3

Report for DKFZphtes3_14g5.3

[LENGTH] 379
 [MW] 43634.03
 [pI] 9.59
 [HOMOL] PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
 [BLOCKS] BL00603D Thymidine kinase cellular-type proteins
 [BLOCKS] BL00530C
 [PROSITE] ATP_GTP_A 1
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 18.73 %

```

SEQ  MVFFTCNACGESVKKIQVEKHVSVCRNCECLSCIDCGKDFWGDDYKNHVKCISEDQKYGG
SEG  .....
PRD  cccccccccccccchhhhhhhhecccccccccccccccccccccccccccccccccccccc

SEQ  KGYEGKTHKGDIKQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhccccchhhhhhhhhhhc

SEQ  SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRLHPVANPHAISTKVPASKVKDAVEQ
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhccchhhhhhhhhccccccccccccccccccccchhhhh

SEQ  QGEVKKNKREERQKKRKREKKELKLENHQENSRNQPKPKRKKGQEADLEAGGEEVPE
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhchhhhhcccccc

SEQ  ANGSAKRSKSKKQKDSASEEARVGAGKRKRHSEVETDSKKKMKLFEHPEGGEPE
SEG  .....
PRD  cccccchhhhhhhhhccchhhhhhhhhccccccccccccccccchhhhhcccccccccc

SEQ  DEAPAKGKFNWGTIKAILKQAPDNEITIKLRKKVLAQYYTVTDEHHRSEEELLVIFNK
SEG  .....
PRD  cccccceehhhhhhhhhhhccccccchhhhhhhhhhhhhccchhhhhhhhhhhhhhh

SEQ  KISKNPFTKLLKDKVLVK
SEG  .....
PRD  cccccchhhhhhhhhhhccc
  
```

Prosite for DKFZphtes3_14g5.3

PS00017 60->68 ATP_GTP_A PDOC00017

(No Pfam data available for DKFZphtes3_14g5.3)

DKFZphtes3_14h21

group: nucleic acid management

DKFZphtes3_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNAtg

Sequenced by BMFZ

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGAA CAATGTCCCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GGTCGTTGCT AGTCGGCGAA GCTCGACAGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGATA TAGTGTGCGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
201 CCCCCGGAGG CCGTGCCCGC TGGTCACGAG GAACTGCCGC TGTGTTTTCG
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTCTGGT GGGTCAAAAA
301 TAAAGAATAT ACAAAGTACA ACAAACACCA CAATCCAAT AATACAAGAA
351 CAACCAGAAAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCAAAACGAA
401 AGCAAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAAATTACA
451 ATTCAGAATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
501 GGAAGCACAG ATAACAATGT TGTTCAGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAAT AGAGAGGAAG GTTGAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTTATA AAGAGTCCAC TGCCACAAGT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTAATAT
701 AACGTGGGAT GACTTGAAGG ATGGGGAGAA ACGACCTATC CCCAATCCTA
751 CCGTGCATTG TGATGACGCC TTTCAATGTT ATCCTGAGGT TATGGAAAAAC
801 ATTAATAAGG CAGGTTTCA AAAGCCAACA CCTATTCACT CACAGGCATG
851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATGTGTGTTT TTAATGCCTG GATTTATTCA TCTGGTCCTT
951 CAACCCAGCC TTAAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAAATG TGCAAAATAT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGTAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTGCTCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTTAGAT GTGCGCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCCTCATTC AGTTCATCGC CTCGCACAAT
1351 CTTATTGTAA AGAACAATG ATTGCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAGTT CAGTGAAGCA AAATATAATT GTAACCACCG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAACAGGCG AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGGTGT TCCATTACAA
1801 CTTTGACTAG AATGATTGG AGGGTTGCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAAT GGAAGAAAAA ATGGAAAGAC
1951 CTCGAAGGAG GCCCAAGAAG TTTCAATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTTTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGAATGATT GACTGATTCT TAAATAATA
2101 GTGTTTGAAA ATATAGAATC CAGTGTTTTA TACTTTCTTT AATAAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648
 Category: strong similarity to known protein
 Classification: Nucleic acid management
 Prosite motifs: ATP_GTP_A (286-294)
 DEAD_ATP_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTTNTT
101 IQIIQEQPES LVKIFGSKAM QTKAKAVIDN FVKKLEENYN SECIDTAFQ
151 PSVGKDGSTD NNVVAGDRPL IDWDQIREEG LKWQKTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNTWDDLKD GEKRPIPNPT CTFDDAFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTGT GKTLCYLMGP
301 FIHLVLQPSL KGQRNRPGML VLTPTRELAL QVEGECCKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDIII ATPGRLNDLQ MSNFVNLKNI TYLVLDEADK
401 MLDMGFEPQI MKILLDVRPD RQTVMTSATW PHSVHRLAQS YLKEPMIVYV
451 GTLDLVAVSS VKQNIIVTTE EEKWSHMQTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQDRREKAL ENFKTGKVRT LIATDLASRG
551 LDVHDVTHVY NFDFPRNIEE YVHRIGRTGR AGRTGSVITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRPKKFH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14h21, frame 3

TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; *Caenorhabditis elegans* cosmid
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like
 protein."; *S.pombe* chromosome II pl p8B7., N = 1, Score = 971, P =
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (*Saccharomyces cerevisiae*), N = 1,
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (*Schizosaccharomyces*
pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; *Caenorhabditis elegans* cosmid
 Y54G11A
 Length = 504

HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREEGLKWQKTKWADLPPIKKNFYKESTATSAMSKEADSWRKENFNTWDDLKDGEK 233
      D+++E  W K      P I ++ YK      +S +      + ++
Sbjct: 23 DRLKDNFSWMK-----PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQIDLIGVAQTGTGKT 293
      IP P  +F+ AF      +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASIMGEIRKNGFEKPSPIQSQMWPLLLSGQDCIGVSQTGSGKT 135

Query: 294 LCYLMPGFIHLVLQPSL-----KGQRNRPGMLVLTPTRELALQVEGECCKYSYKGLRSVC 348
      L +L+P +H+ Q +      + Q+ P +LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALLHIDAQLAQYEKNDDEEQKPSFFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGNRDEQIEELKKGVDIIATPGRLNDLQMSNFVNLKNITYLVLDEADKMLDMGFEP 408

```



```

+YGGG+R EQ+E + GV+I+IATPGRL DL ++L ++TY+VLDEAD+MLDMGFE
Sbjct: 196 LYGGGSRPEQVEACRGGVEIIVATPGRLTDLSDGVISLASVTYVVLDEADRLDMGFEV 255
Query: 409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVT 468
      I +IL ++RPDR +TSATWP V +L Y KE ++ G+LDL + SV Q
Sbjct: 256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKEAVMAVNGSLDLTCKSVTQFFEFV 315
Query: 469 TEEKW---SHMQTFLQSMSSTD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
      + ++ + FL + + K+I+V K +ADHLSSD + I+ + LHG R Q
Sbjct: 316 PHDSRFLRVCEIVNFLTAAHQNYKMIIFVKSKVMADHLSSDFCMKGINSQGLHGGRSQS 375
Query: 525 DREKALENFKTGKVRILIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRT 584
      DRE +L ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct: 376 DREMSLNMLRSGEVQILVATDLASRGIDVDPDITHVLNVDFFMDIEEYVHRVGRTRAGR 435
Query: 585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMARFEAHQRKREMERKMERPOGRP 644
      G +++ L ND LI ILE++ Q +P++L AE++ K + R RP R
Sbjct: 436 GEAMSLFWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KCQSGROGPRPSFRN 492
Query: 645 KK 646
      K
Sbjct: 493 NK 494

```

Pedant information for DKFZphtes3_14h21, frame 3

Report for DKFZphtes3_14h21.3

```

[LENGTH]      648
[MW]           72873.51
[pI]           8.84
[MOMOL]        TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-
101
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       04.05.03 rna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-70
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
[FUNCAT]       1 genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 2e-49
[FUNCAT]       j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-48
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 1e-07
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 4e-96
[PIRKW]        RNA binding 3e-87
[PIRKW]        DEAD box 5e-50
[PIRKW]        transmembrane protein 4e-27
[PIRKW]        DNA binding 3e-67
[PIRKW]        recF recombination pathway 3e-10
[PIRKW]        ATP 4e-96
[PIRKW]        purine nucleotide binding 5e-50
[PIRKW]        P-loop 4e-96
[PIRKW]        hydrolase 9e-45
[PIRKW]        protein biosynthesis 5e-50
[PIRKW]        ATP binding 1e-61
[SUPFAM]       WW repeat homology 8e-88
[SUPFAM]       DEAD/H box helicase homology 4e-96
[SUPFAM]       unassigned DEAD/H box helicases 7e-87
[SUPFAM]       ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM]       ATP-dependent RNA helicase DHH1 2e-43
[SUPFAM]       recQ protein 3e-10
[SUPFAM]       Bloom's syndrome helicase 5e-07
[SUPFAM]       translation initiation factor eIF-4A 5e-50
[SUPFAM]       recQ helicase homology 3e-10
[SUPFAM]       tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE]      DEAD_ATP_HELICASE 1

```


[PROSITE] ATP_GTP_A 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] KH domain family of RNA binding proteins
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 8.49 %

```

SEQ  MSHHGGAPKASTWVVASRSSTVSRAPERRPAEELNRTGPEGYSVGRGGRWGTSRPPEA
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  cccccccccceeeeeccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VAAGHEELPLCFALKSHFVGAIVGRGGSIKNIQSTNTTTIQQEQPESLVKIFGSKAM
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  cccccccccchhhhhccccceeeccccccccccccccccccccceeeccccccccccccchh

SEQ  QTKAKAVIDNFVKKLEENYNSECGIDTAFQPSVGKDGSTDNVAVAGDRPLIDWDQIREEG
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc

SEQ  LKWQKTKWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGKRPINPT
SEG  .....
PRD  chhhhhhhccccccccccccccccccccchhhhhhhhhhhhhheeeccccccccccccccccc

SEQ  CTFDADFQCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQIGIDLIGVAQTGTGKTLCYLMPG
SEG  .....
PRD  cccccccccchhhhhhhhhhhccccccccccccccccccccceeeccccccccceeeccce

SEQ  FIHLVLQPSLKGQRNRPGLVLTPTRELALQVEGECKYSYKGLRSVCVYGGGNRDEQIE
SEG  .....
PRD  eeeccccccccccccceeeccccchhhhhhhhhhhhhhhccccceeeccccccccchhhh

SEQ  ELKKGVDIIATPGRLNDLQMSNFVNLKNITYLVLEADKMLDMGFEPQIMKILLDVRPD
SEG  .....
PRD  hhhheeeccccccccchhhhhhhccccceeehhhhhhhhccccchhhhhhhhhhhccc

SEQ  RQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
SEG  .....
PRD  ceeeccccchhhhhhhhhhhheeeccccccccccccceehhhhhchhhhhhhhh

SEQ  LQSMSSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTGKVRI
SEG  .....
PRD  hhhheccccceeehhhhhhhhhhhhhhccccceccccchhhhhhhhhhhcccce

SEQ  LIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  eeehhhhhccccccccceeeccccccccceeeccccccccceeeccccchhhh

SEQ  ELINILERANQSIPEELVSMARFEAHQRKREMERKMERPQGRPKKFH
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  hhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccc

```

Prosites for DKFZphtes3_14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphtes3_14h21.3

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQaIPiILeGRDVMACaQTGSGKTAaF P++++NI+++GF KPTPIQ+QA+PI+L+G D+++ AQTG+GKT+++		
Query	248	QCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQIGIDLIGVAQTGTGKTLCY	296
HMM	1IPMLQHIDwdPWpppPQd..PrALILAPTRELAMQIEEcRkFgkHMng L+P ++H+ +P +++ Q+ P +L+L+PTRELA+Q++ EC K+++ +		
Query	297	LMPGFIHLVLQP-SLKGQRNRPGLVLTPTRELALQVEGECKYSYK-G-	343
HMM	IRImciYGGtnMRdQMRmLeRGpPHIVATPGRLIDHIERgtldLDriEm +R++C+YGG N ++Q+++L++G+ +I+IATPGRL D+ +++ ++L++I++		
Query	344	LRSVCVYGGGNRDEQIEELKKGv-DIIATPGRLNDLQMSNFVNLKNITY	392
HMM	LVMDEADRMMDMGFIDQIRrImrqIPmpwNRQTMFSATMPdeIqELARr LV+DEAD+MLDMGF++QI++I+ ++ ++RQT+M SAT+P ++ +LA		
Query	393	LVLDEADKMLDMGFEPQIMKILLDVR--PDRQTVMTSATWPHSVHRLAQs	440


```

HMM          FMRNPiRInId.MdELttnEnIkQwYiyVerEMWKfdClerLie*
++++P + ++ D +++ +KQ +I+ E++K + ++++
Query        441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-EEEKWSHMQTFLQ 482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
+ + ++++G++IG+GGS I++I++ ++++I I++E+ + + + I
Query        71 CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIIEQ-P---ESLVKIF 115

HMM          G*
G
Query        116 G 116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWLknl....GirvmYIHGdMpQeERdeIMddFNnGEynVLicTD
+ +++ L+ + +I+V ++HGD++Q++R++++++F++G+ ++LI+TD
Query        497 KAVADHLSSDLILGNISVESLHGDREQDREKALENFKTGKVRILIATD 545

HMM          VggRGIDIPdVNVHVINYDMPWNPEqYIQRIGRTgRIG*
+++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query        546 LASRGLDVHVDVTHVYNFDFPRNIEEYVHRIGRTGRAG 582

```


DKFZphtes3_14p14

group: testes derived

DKFZphtes3_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```
1 GAAGCCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTC
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTG
101 GGTGCTGAAG GCCAGGCAGA GCATTGCGCT GTAGGGAGGC CGATCCTCCT
151 CGGGCCTGTT ACCGGCGGGT CTTTGTCTTT AGACCTGGGG TTCTTGGCCT
201 CACGGATTCC AAGGAATGGA ACGTTGGGCC ATGCGTGTGA ACGAGCTCTA
251 TGTCTGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGTCTGA
301 ACATCAGTTT ACCCAATCTG CACTGCCAGT TGGTTGGGCT TGACATTGAG
351 GATGAGATGG GCAGGCACGA AGTGGGCCAC ATCGACAAC CCATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGCCGCTT CGAGGGGCGA TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTTCCT
501 GCTGCTCTCC CGCTCCCTGG AAATGGTTG TGGAGGCACT CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TTCGTGTCCA GCTGCCTGGC
601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTAAAG
651 CCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTGTCAG CCCCAGCCCC TGGGTTCAAG
751 TCCAGCTCT ACCCCTTCTT GGCCCTTACA AGTCACTTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
851 AGTTTGTGTG GGGATTAAAT GAGCTAATGC AGATTCAATC ATTCAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CACGGTTCCT GGGCTGGTGG GGCTCCCATC CTGGTAGAGG GAGACAGTCT
1001 ACAAAACAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGGT GGGCTTGAAG GTACCTTGTC CGCTCAGAAG
1101 GACCAATCAA GGTTCATCTG TGTGTTGTCC TCAGAACCAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCGGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACGCTGTGAG GCAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGCCAGAGA GGTAAAGTGT CTTGCCCGAG
1351 GTACACAGC TATCTGCATG TCCCAACAAC CCCCTTCCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCTCT CAGACCCCTGA CACCTGAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAGATAA ATTTGTTTGG TCCAGTATAG
1551 CATTTAACTG CATCAGAACC AGTATGAAAA GACCAGGAAT CCAGATTCTT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAG TGGGCAGTGT TCCCCTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCTGG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGCTAAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGCT
1851 GGGAGAAATC TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACCAGCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCGTG
1951 GTTCACTGGG GGCTCTCTCT GCCCATGAGC TGCCACACAG CACCTTGGCC
2001 ATGCCCCGTA ATTTGGATT TATGGTGGTT GTGATGGAAG GCCATTGAG
2051 GGTGTTGAAC AGGGAGGCAA TGTAATCAGA TTTATGCCTT AGAACTGGAC
2101 TATCCAATAG GTTGCCACCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTATC AATTCACTCA CTCATTCTCA TCAACCACAT TTCAGTGCT
2201 CAAAGCCAGC TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTCTGCTG GGACCGCACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCAGCAGG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCCTCTACT AAAAATACAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCGTAC CACTGCACCT CAGTCTGGGC GACAGAGTGA GACTCCATCT
2601 CAAAACAAAC AAAAAAAGGA TGGGGCTGGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTGTGGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGGG CACATGTGGA
```



```

2751 AGCAAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCCCTGTTT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACTG GTCATTCCAC TCTGGTTTCC TTTGGGGTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAGTTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTTCTGG ACCACAGATG TTGAGGTGG AGTCATTAGT GAATAGATGA
3001 TGTTGGAAAGT CATGGGTCCT CAGAGTGGGG GCTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTGTCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTCC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG
3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCCAGCAC TTTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCTG GTCACATGT
3601 CAAAACCTG TATCTACAAA AAAATACAAA AGTTAACCAA GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGTCTT CTGAGCCCAG
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAACCCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATTT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTAA AATATTTTAA TTACTATTTA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159
 Category: putative protein
 Classification: no clue

```

1 MERWAMRVNE LYVDDPDKDS GKKIDVSLNI SLPNLHCELV GLDIQDEMGR
51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPLSPFY LLPFAVSPL
101 PGNWLWRHSL DLTLTQPPAS EGSCPAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSLQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p14, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_14p14, frame 3

Report for DKFZphtes3_14p14.3

```

[LENGTH]      159
[MW]           17778.55
[pI]           5.74
[FUNCAT]      99 unclassified proteins    [S. cerevisiae, YAL042w] 5e-04
[KW]          Alpha_Beta

```

```

SEQ  MERWAMRVNELYVDDPDKDSGGKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM
PRD  cccchhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  KIPLNNGAGCRFEGQFSINKVWKPLSPFYLLPFAVSPLPGNWLWRHSLDLTLTQPPAS
PRD  eecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```


SEQ EGSCPAAWPFLRLRIWMGVQAPWGFKPLMAGSGRSYSSLQ
PRD cccccchhhhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3_14p14.3)

(No Pfam data available for DKFZphtes3_14p14.3)

DKFZphtes3_14p7

group: testes derived

DKFZphtes3_14p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1 GGAATCCCAA GAAACAGTTA TGATGGGGGA CTCTATGGTG AAAATAAATG
51 GGATTTATTT AACAAAATCA AATGCTATTT GCCACTTAAA GAGTCACCCA
101 CTTTCAGTAA CTGATGATGG AGGCTTCAGT GAAATAAAGG AGCAAGAAAT
151 GTTCAAAGGA ACAACATCTT TACCATCTCA TCTCAAGAA GGAGGGGACC
201 AGGGGAAGAG ACATGCGAGG GCCTCATCAT GCCCCAGTAG CTCAGACCTG
251 AGCAGGCTGC AAACCAAAGC AGTCCCAAAA GCTGACCTGC AAGAAGAGGA
301 CGCAGAAATA GAAGTAGACG AAGTCTTTTG GAATACAAGG ATTGTACCGA
351 TTTTGCGTGA ATTAGAAAAG GAAGAAAACA TGAACACGGT TTGTGCTGCT
401 TGCACACAAC TTCATCATGC TTTAGAGGAA GGAAACATGC TTGGAATATA
451 ATTTAAGGGA AGAAGTATTC TCCTGAAGAC CCTGTGTAAA CTAGTTGATG
501 TTGTTTCAGA CTCGCTCAGC CTTAAACTTG CAAAATAAAT TCTAGCACTT
551 AAAGTGAGTA GAAAGAATCT TCTTAATGTC TGCAAACTTA TATTTAAAT
601 TAGCAGGAAT GAGAAGATG ATTCTTTGAT TCAAAATGAC AGCATTCTGG
651 AATCATTATT GGAGGTACTA AGAAGTGAAG ACCTGCAAAC TAACATGGAA
701 GCTTTTTTAT ACTGTATGGG GTCTATAAAG TTCATTCTG GAAATCTGGG
751 ATTTCTTAAT GAAATGATCA GCAAAGGTGC TGTGGAATA CTGATAAAT
801 TGATAAAACA AATAAATGAG AACATCAAGA AATGTGGTAC ATTTTTCGCT
851 AATTCGGGCC ACTTGCTAGT CCAGGTGACT GCTACATTGA GAAACTTGGT
901 TGATTATCAT TTAGTAAGAA GTAAGTTCCT AAACATCAGT GCCCTTCCCC
951 AGCTCTGCAC GGCAATGGAA CAGTACAAGG GTGACAAGGA CGTCTGTACC
1001 AATATTGCCA GAATATTCAG CAAACTTACT TCTTACCGTG ACTGCTGCAC
1051 AGCCTTGGCC AGCTATTCCA GATGTTATGC CTTATTCTG AATCTAATTA
1101 ACAATACCA GAAGAAGCAG GATTTAGTCG TCCGTGTTGT TTTTATTCTT
1151 GGCACCTTGA CGGCAAAAAA TAACCAGGCT CGTGAACAA TTTCCAAAGA
1201 GAAAGGGAGC ATCCAACTC TGCTGTCAAT ATTCCAGACG TTCCATCAGC
1251 TGGATCTGCA TTCCAGAAAG CCGGTGGGCC AACGAGGCGA GCAGCAGAGG
1301 GCGCAGAGGC CGCCGTCAGA GGCAGAGGAC GTGCTCATCA AGCTGACTCG
1351 TGTGCTGGCC AACATTGCCA TCCACCCGGG CGTGGGCCCG GTGCTGGCCG
1401 CCAACCCGGG GATAGTGGGC CTGCTCCTGA CCACGCTGGA ATACAAGTCA
1451 CTTGATGATT GTGAGGAGCT GGTGATCAAT GCTACAGCGA CAATCAACAA
1501 TTTATCTTAC TACCAAGTGA AGAATTCCAT AATTCAAGAC AAAAAGCTAT
1551 ATATTGCTGA ATTGCTCTTA AAGCTTCTTG TCAGTAACAA CATGGATGGA
1601 ATCCTGGAGG CTGTGCGTGT TTTCCGAAAT CTCTCCAGG ACCATGATGT
1651 CTGCGATTTC ATTGTGCAGA ACAATGTCCA CAGGTTTCATG ATGGCGCTGC
1701 TGGATGCTCA GCATCAGGAT ATCTGCTTTT CTGCGTGTGG TGTTCCTCTC
1751 AATCTCACTG TGGATAAAGA CAAGCGTGTC ATCTTGAAAG AAGGAGGTGG
1801 CATTAAAAAG TTAGTGGACT GTTAAAGAGA TTTGGGTCCT ACTGATTGGC
1851 AGCTGGCCTG CTGGTTTGT AAAACTTTAT GGAACCTCAG TGAAAACATC
1901 ACTAATGCTT CGTCATGTTT TGGAAATGAA GACACCAACA CACTCTTACT
1951 CTTGCTCTCA TCATTTTATG ATGAAGAACT AGCACTGGAT GGCAGTTTGT
2001 ATCCAGACCT AAAAACTAT CACAACTCC ATTGGGAAAC AGAATTCAAA
2051 CCTGTGGCAC AGCAGCTTCT AAACCGAATT CAGAGACATC ACACCTTCCT
2101 GGAACCCCTG CCCATTCCCT CTTTCTAACA TGATGCAGAT TAACAGTAGA
2151 AACGAGAACT CACGCTCTCC TCATTCTTAA GAACTGGTAA CAAACGTGAA
2201 CATTTTTTTC AGCATTAAAC AATGTGGAAA GTTTTTCAAG AACTGGTTTT
2251 AGTGAGTAGC TGAAGTATTT TTTAAATTA AGCATTCTCT CTTGTAGGT
2301 ATTATGAAAA AATGAATATA CACATTATAT TTCCTGTGTA GAGAAATGTA
2351 AGATGAAAAA ATGTGCATTT TCAAGTAAAT GACTTTTTCT TCTATTCTCT
2401 ATTAACAAT TTAGTTCTAG TCTTAAAAA AAAAAAATA AAAAAAATA
2451 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702
Category: putative protein

```

1 MMGDSMVKIN GIYLTksNAI CHLkSHPLQL TDDGGFSEIK EQEMFKGTTs
51 LPSHLKNGGD QGKRHARASS CPSSSDLSRL QTKAVPKADL QEEDAEIEVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFKGRSI
151 LLKTLCKLVD VGSDSLskKL AKIILALKVS RKNLLNVCKL IFKISRNEKN
201 DSLIQNDsIL ESLLVLRSE DLQTNMEAFI YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIKQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSSSLVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RCYALFLNLI NKYQKKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQT
401 LLSLFQTFHQ LDLHSQKPVG QRGEQRAQR PPSEAEVDLI KLTRVLANIA
451 IHFGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSYYQV
501 KNSIIQDKKL YIAELLKLL VSNMMDGILE AVRVEGNLSQ DHQVCDFIVQ
551 NNVHRFMMAI LDAQHQDICF SACGVLLNLT VDKDKRVILK EGGGIKKLVD
601 CLRD LGPTDW QLACLVCKTL WNFSENITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFPKVA QQLLNRIQRH HTFLEPLPIP
701 SF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p7, frame 2

TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,
complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete
cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 45/163 (27%), Positives = 77/163 (47%)

```

Query:  442 LTRVLANIAIHFGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSYYQVK 501
      L +++ NI+ H G P          VG L + S D+ EE VI  T+ NL+  +
Sbjct:  483 LMKMIRNISQHDG--PTKNLFIDYVGDLAAQI---SSDEEEEFVIECLGTLANLTIPOLD 537

```

```

Query:  502 -NSIIQDKKLYIAELLKLLVSNMMDG-ILEAVRVFGNLSQDHQVCDFIVQNNVHRFMMA 559
      ++++ KL + L KL      D +LE V + G +S D  + ++ +  ++
Sbjct:  538 WELVLKEYKL-VFPLKDKLPGAEDDLVLEVVMIGTVSMDDSCAALLAKSGIIPALIE 596

```

```

Query:  560 LLDQHQDICFSACGVLL---NLTVDKDKR-VILKEGGGIKKLVDCIRD 604
      LL+AQ +D F C ++      +  + R VI+KE      L+D + D
Sbjct:  597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIKETQAPAYLIDLMDH 644

```

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 42/178 (23%), Positives = 82/178 (46%)

```

Query:  169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDsILESLLVLRSEDLQTNME 227
      K K      L V ++ LL V  L+  ++ + + +  ++N +I+ L++ L +  N E
Sbjct:  263 KTFKKYQGLVVQEQELLRVALYLLLNLAE DTRTELKMRNKNIVHMLVKALDRD----NFE 318

```

```

Query:  228 AFLYCMGSIKFISGNLGFLEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVT 287
      + + +K +S +  N+M+  VE L+ +I  +E++      L + +
Sbjct:  319 LLILVVSFLKKSIFMENKNMDMVEMDIVEKLVKMIPCEHEDL-----LNITLR 366

```

```

Query:  288 ATLRNLVDSSSLVRSKFLNISALPQLCTAM--EQYKGDQDVCT--NIARI--FSKLTSYRD 341
      L      D+ L R+K + +  LP+L  + E YK  +C  +I+  F  + +Y D
Sbjct:  367 LLLNLSFD TGL-RNMVQVGLLPKLTALLGNENYK-QIAMCVLYHISMDDRFSKSMFAYTD 424

```


Query: 342 CCTAL 346
C L
Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAELLKLLVSNMMDGILEAVRVFGNLSQDHDVCDFFIVQNNVHRFMMALLDAQHQDICFS 571
I +L+K L +N + ++ V LS + + +V+ ++ ++ ++ +H+D+
Sbjct: 304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDVMEMDIVEKLVKMIPCEHEDLLNI 363

Query: 572 ACGVLLNLTVDKDRVILKEGGGIKKLVDCRLDGLPTDW-QLACLVCKTLWNFSENITNA 630
+LLNL+ D R + + G + KL L G ++ Q+A +C L++ S +
Sbjct: 364 TLRLLLNLSFDTGLRNKMVQVGLLPKLTALL---GNENYKQIA--MC-VLYHISMD-DRF 416

Query: 631 SSCFGNEDT-NTLLLLSSFLDEELALD 657
S F D L+ +L DE + L+
Sbjct: 417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 LIFKISRNEKN-DSLIQNDISLESLLLEVLSE-----DLQTNMEAFLYCMGSIKIFISG 241
LI +++RN N + L+ N++ L +L VLR + +L TN+ +C S G
Sbjct: 155 LILQLARNPDNLELLLNELTALGALARVLRDQKQSVELATNIIYIFFCFSSFSHFHG 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 LVQVTATLRNL---VDSSLVRSKFLNISALPQLCTAMEQYKGDQVCTNIARIFSKLTS 338
+++ TL NL +D LV ++ +P L ++ + D+ + I S
Sbjct: 521 VIECLGTLANLTIPDLWELVLKEY---KLVFPLKDKLPGAAEDDLVLEV-IMIGTVS 576

Query: 339 YRDCCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKERGSI 398
D C AL + S + L+N Q+ + V +++++ + + R+ KE +
Sbjct: 577 MDDSCAALLAKSGIIPALIELLNAQQEDEFVQCIIYVFYQMFV-HQATRDVVIKETQAP 635

Query: 399 QTLSSL 404
L+ L
Sbjct: 636 AYLIDL 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKKLYIAELLKLLVSNMMDGILEAVRVFGN 537
CE E ++N T + NLS+ ++N ++Q + + L LL + N I A+ V +
Sbjct: 355 CEHEDLLNITLRLLLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query: 538 LSQDHDVCD-FIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDRVILKEGGGIK 596
+S D F + + + M L + + I +NL +K ++ EG G+K
Sbjct: 410 ISMDDRFKSMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVLICEGNGLK 469

Query: 597 KLVDCRLDGLPTDWQLACLVCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEELAL 656
L+ R L D L+ K + N S++ + F + L +SS +EE +
Sbjct: 470 MLMK--RALKLKD---PLLMKMRNISQHDGPTKNLF-IDYVGDLAAQISSDEEEFVI 522

Query: 657 D 657
+
Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 LNISALPQLCTAM-EQYKGDQVCTNIARIFSKLTSYRDCCTALASYSRCYALFLNLINK 362
LN +AL L + E +K ++ TNI IF +S+ + Y + AL +N+I+
Sbjct: 171 LNETALGALARVLRDQKQSVELATNIIYIFFCFSSFSHFHGLITHY-KIGALCMNIIDH 229

Query: 363 YQKKQDL 369
K+ +L
Sbjct: 230 ELKRHEL 236

Pedant information for DKFZphtes3_14p7, frame 2

Report for DKFZphtes3_14p7.2

[LENGTH] 708
[MW] 79266.35
[PI] 6.57


```

[FUNCAT]      30.25 vacuolar and lysosomal organization      [S. cerevisiae, YEL013w] 3e-04
[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w]
3e-04
[FUNCAT]      09.25 vacuolar and lysosomal biogenesis      [S. cerevisiae, YEL013w] 3e-04
[BLOCKS]      BL00923F Aspartate and glutamate racemases proteins
[BLOCKS]      BL00288B Tissue inhibitors of metalloproteinases proteins
[PROSITE]      MYRISTYL          9
[PROSITE]      AMIDATION         1
[PROSITE]      CK2_PHOSPHO_SITE   12
[PROSITE]      PKC_PHOSPHO_SITE   7
[PROSITE]      ASN_GLYCOSYLATION  11
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY     7.49 %

```

```

SEQ  ESKETVMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFKGTTSLPSH
SEG  .....
PRD  cccceeeccccceccccccccceeeccccccccccccccccchhhhhhhcccccccccc

```

```

SEQ  LKNGGDQGRHARASSCPSSDLSRLQTKAVPKADLQEEAEIEVDEVFWNTRIVPILRE
SEG  .....xxxxxxxxx.....
PRD  cccccccchhhhhccccccccchhhhhhhccccchhhhhhhhhhhccccceehhhhh

```

```

SEQ  LEKEENIETVCAACTQLHHHALEEGNMLGNKFKGRSILLKTLCKLVDSLSLKLAKII
SEG  .....xxxxxxxxx.....
PRD  hhhhccccchhhhhhhhhhhhhccccccccccccccccchhhhhheeeccccchhhhhhh

```

```

SEQ  LALKVSRKNLLNVCKLIFKISRNEKNDSLIQNDSILESLLVLRSEDLTQNMEEFLYCMG
SEG  xxxx.....
PRD  hhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccccchhhhhhhhhcc

```

```

SEQ  SIKFISGNLGFLENMISKGAVEILINLIKQINENIKKCGTFLPNSGHLVQVTATLRNLV
SEG  .....
PRD  ceececcccchhhhhhhccccchhhhhhhhhhhccccccccccccceeeehhhhhhhhh

```

```

SEQ  DSSLVRSKFLNISALPQLCTAMEQYKGDVDCTNIARIFSKLTSYRDCCTALASYSRCYA
SEG  .....
PRD  ccchhhhhheeeccccchhhhhhhhhccccceeeehhhhhhhhhccccchhhhhhhhhhh

```

```

SEQ  LFLNLINKYQKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSITQLLSLFQTFHQDLH
SEG  .....
PRD  hhhhhhhhhhhhhhhheeeccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhcc

```

```

SEQ  SQKPVGQRGEQHRAPPPSEADVLIKLTRLANIAIHGPGVPVLAANPGIVGLLLTTLE
SEG  .....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhccccccccceeeccccchhhhhhhhh

```

```

SEQ  YKSDDCEELVINATATINNLSYYQVRNSIIQDKKLYIAELLKLLVSNMMDGILEAVRV
SEG  .....xxxxxxxxx.....
PRD  hhccccchhhhhhhheeeccccccccceeeehhhhhhhhhhhhhccccchhhhhhhhh

```

```

SEQ  FGNLSQDHDVDCFIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGG
SEG  .....
PRD  cccccccccceeeccccchhhhhhhhhhhccccceeeccccceeeccccceeecccccc

```

```

SEQ  IKKLVDCRLDGLPTDWQLACLVCRTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEEL
SEG  .....xxxxxxxxx.....
PRD  hhhhhhhhhccccccccchhhhhhhhhccccccccccccccccccccceehhhhhhhhh

```

```

SEQ  ALDGSFDPDLKNYHKLHWETEFKPAQQLLNRIQRHHTFLEPLPIPSF
SEG  xxx.....
PRD  hhccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhheeecccccc

```

Prosite for DKF2phtes3_14p7.2

PS00001	206->210	ASN_GLYCOSYLATION	PDOC00001
PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00001	311->315	ASN_GLYCOSYLATION	PDOC00001
PS00001	385->389	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00001	500->504	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	584->588	ASN_GLYCOSYLATION	PDOC00001
PS00001	628->632	ASN_GLYCOSYLATION	PDOC00001
PS00001	632->636	ASN_GLYCOSYLATION	PDOC00001
PS00001	635->639	ASN_GLYCOSYLATION	PDOC00001
PS00005	173->176	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_14p7.2)

DKFZphtes3_15a13

group: testes derived

DKFZphtes3_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp.116, 3 EST hits

S.cerevisiae Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAAGCGCA  TCGCGCTCGG  GCACAGCGCG  TGCAGCCTCG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCCGC  TTGCAGAGGA  CTCCCATGAG
151  TGCAGTGGTA  TTTCCCAATA  AGATATCAAC  TGAACACCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTGAGG
251  GGAATATTCC  CAGAAATGCG  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGTCAAAATA  CTGAGAGAAG  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAAAACCAG  AAGATCCTCA  GACAAATTTCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTTCATAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTTG  TCTACTGACA  CCAAGAAAGC  AAGCATCTCT
551  CTCATTCGCA  AGATTATAT  CCTAATGCAA  AATCTGGGGC  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGITATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAATA
801  TTGACTCAAC  TATACTATCA  CCAAACCAAA  TAAACACACC  ATTTCAAAAA
851  ATCCTGAGGG  ACAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTTGGAC  ATTGAACTA  AAATGGAAGA  ACAGGAAAAA  AACCCTGCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAATTATG
1001  AGGTCTAAAG  AAAGTCCAGA  TCTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAATAGT
1201  CCTCCATCAC  TTTGATTCTT  CTAGTCAAGA  GTCAGTGCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTTGCAGAGT  TCTTCTCACC  ATTTAACTG  AAGGACCCTA  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACACTAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATAA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTC
1501  GGTAAATAAGT  AAAATTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCCTTTAAA  TATATACTAA  CTGTTAAGGC  CCTCTAATGC  CATTTTCTTA
1601  AACAGTAATG  TTTACTTTGG  TATTAAAAAT  TGGTATGGAT  TCACCTTTTA
1651  CTTATGTTAA  AATTATACCA  TTTAACTGGC  TCTTTTGTC  TTGTGCTGTT
1701  ATTAACAACA  TGTTCTTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAGG
1801  GCGGCCGCGT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387
Category: similarity to known protein

```

1 MATAQLQRTP MSALVFPNKI STEHQSLVLV KRLAVSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDE NCPGSTQLVK WMLGCDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNGPL MDFISKQSN ESSMLSTDTK KASILLIRKI
151 YILMQLGLPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFVKVT TERERENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLISHSQV EQLVNKTSSEL DMSESKTRSG KVFQNMANG NQPVKSSKEN
351 RKRSHESGR IVLHHFSSS QESVPKRRKF SEPKEHI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15a13, frame 2

TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.
Length = 562

HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22
Identities = 84/290 (28%), Positives = 145/290 (50%)

```

Query:   22 TEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSTQLVKW 81
          TE SL+L + LL +++ I+Y+RG+FPE + + + L +KI + S +L+ W
Sbjct:   11 TEQDSLLLRNLLRIAIFNISYIRGLFPEKYFNDKSVPALDMKIKLMPMDAESRRLIDW 70

Query:   82 M-LGCDALQKKYVYT-----NPEDPQTISECYQFKFKYTNGP--LMDFISK--NQSN 130
          M G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct:   71 MEKGVYDALQRKYLKTLMFISICETVDGPMIEE-YSFYSYSDSDSQDVMNINRTGNKKN 129

Query:   131 ESSMLST-----DTKKASILLIRKIYILMQLGLPLPNDVCLTMKLFYYDEVTPPDYQPP 184
          ST + + + +R + LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct:   130 GGIFNSTADITPNQMRSSACKMVRTLVQLMRTLDMKPDERTIVMKLLYYDDVTPPDYEP 189

Query:   185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFVKVTT-----ERERENIDSTILS 235
          F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct:   190 FFRGCTEDEAQYVWTKNPLRMEIGNVNSKHLVLTLLKVKSVLDPCEDENDMQD-DGKSIG 248

Query:   236 PKQIKTPFQKILRDKDVEDEQEHY-----TSDDLDIETKMEEQKNPASSE 281
          P + Q D ++ QE+ DD D E ++ ++PA +E
Sbjct:   249 PDSVHDD-QPSDSDSEISQTQENQFIVAPVEKQDDDDGEVDEDDNTQDPAENE 300

```

Pedant information for DKFZphtes3_15a13, frame 2

Report for DKFZphtes3_15a13.2

```

[LENGTH]      387
[MW]           44417.64
[pI]           5.57
[HOMOL]        TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, Y1072w] 7e-11
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, Y1072w] 7e-11
[FUNCAT]       03.13 meiosis [S. cerevisiae, Y1072w] 7e-11
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, Y1072w] 7e-11
[PIRKW]        nucleus 2e-09
[PIRKW]        zinc finger 2e-09

```


SEQ PRD	MATAQLQRTPMSALVFPNKISTEHQSLVLVKRLLAVSVSCITYLRGIFPECAYGTRYLDD ccccccccccccccccccccchhhhhhhhhhhhhhhhhhhheeeeeccccccccccccch
SEQ PRD	LCVKILREDKNCPGSTQLVKWMLGCDALQKKYVYTNPEDPQTISECYQFKFKYTNNGPL hhhhhhhccccccccccccccccchhhhhhhhhhhccccccccchhhhhheeeeeccccc
SEQ PRD	MDFISKQSNSSMLSTDTKKASILLIRKIYILMQNLGLPLNDVCLTMKLFYYDEVTPPD eeeeccccccccceccccchhhhhhhhhhhhhhhhhccccccccceeeeeeeeeccccc
SEQ PRD	YQPPGFKDGDCEGVIFEPEPMYLNVGEVSTPFHIFVKVVTTERERMENIDSTILSPKQIK ccccccccccccceeeeeccceeeccccccccceeeeeecccchhhhhccccccccchhh
SEQ PRD	TPFQKILRKDVEDEQEHYTSDDDLDIETKMEEQEKNPASSELEEPSLVCEEDEIMRSKES hhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccccccchhhhhhhhhcc
SEQ PRD	PDLSISHSQVEQLVNKTSSELDMSSEKTRSGKVFQNKMGANGPQVKSSKENRKRSQHESGR ccccccchhhhhhhhhhhccccccccccccccccceeeccccccccchhhhhhhhhhhccce
SEQ PRD	IVLHHFDSSSQESVPKRRKFSEPKHEI. eeeecccccccccccccccccccccccccccccccc

PS000001	127->131	ASN_GLYCOSYLATION	PDOC000001
PS000001	130->134	ASN_GLYCOSYLATION	PDOC000001
PS000001	315->319	ASN_GLYCOSYLATION	PDOC000001
PS000004	140->144	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	351->355	CAMP_PHOSPHO_SITE	PDOC000000
PS000004	378->382	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	139->142	PKC_PHOSPHO_SITE	PDOC000005
PS000005	167->170	PKC_PHOSPHO_SITE	PDOC000005
PS000005	221->224	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000005	329->332	PKC_PHOSPHO_SITE	PDOC000005
PS000005	346->349	PKC_PHOSPHO_SITE	PDOC000005
PS000005	358->361	PKC_PHOSPHO_SITE	PDOC000005
PS000006	96->100	CK2_PHOSPHO_SITE	PDOC000006
PS000006	103->107	CK2_PHOSPHO_SITE	PDOC000006
PS000006	177->181	CK2_PHOSPHO_SITE	PDOC000006
PS000006	221->225	CK2_PHOSPHO_SITE	PDOC000006
PS000006	260->264	CK2_PHOSPHO_SITE	PDOC000006
PS000006	268->272	CK2_PHOSPHO_SITE	PDOC000006
PS000006	280->284	CK2_PHOSPHO_SITE	PDOC000006
PS000006	308->312	CK2_PHOSPHO_SITE	PDOC000006
PS000006	318->322	CK2_PHOSPHO_SITE	PDOC000006
PS000006	346->350	CK2_PHOSPHO_SITE	PDOC000006
PS000006	354->358	CK2_PHOSPHO_SITE	PDOC000006
PS000006	369->373	CK2_PHOSPHO_SITE	PDOC000006
PS000008	84->90	MYRISTYL	PDOC000008

580

DKFZphtes3_15c24

group: metabolism

DKFZphtes3_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatgG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```

1 CGAAGGCGGC GGC GAAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGAACT TGCC CAGGAG AGGAGTCTGC AGGTCCCGAG GAGCGGCGAC
151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGG
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATTGTAA
251 GCGACTATGA GAAATCCCGT ACCTTTGCCG TAGCAATAGT AGGTGTTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTGAAG TACACAATA
501 TAAATATAAC ACAGTGGAAA ACTTTCAACA TTTCTAGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTGTA ATGAAC TTGG
651 ACAACATGGA ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGCGATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTCCGCT TGCTCCACCA
751 CTTGTAGTTG CTGCAATAT TGATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTGTGCAGCC AGTCTTCCTA CCACTATGGG TGTGGTTGCT GGGATCTTAG
851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCACTG CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAGAG GAACTGAAAA ATTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATAA ACAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGAAGA TTCTGGTGAA AGCTTGGAAG ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAA TGATAAAACT
1351 TAGGGCAACA TTAATTAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAATCCTG TGACTTGCCT GTTTCTCCCC GCTCCAACGA AATCATTAAAC
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTCTATAAG AAAATGCCCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACTTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAAAATAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAAAA
1951 AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404
 Category: similarity to unknown protein
 Classification: Metabolism
 Prosite motifs: D_2_HYDROXYACID_DH_1 (76-105)

```

1 MAESVERLQQ RVQELERELA QERSLOVPRS GDGGGGRVRI EKMSSEVVDS
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LFPDYDKVEL ANMNRLEFFQ HQAGLSKVQA AEHTLRNINP DVLFEVHNYN
151 ITTVENFQHF MDRISNGGLE EGKPVLDVLS CVDNFEARMT INTACNELGQ
201 TWMESGVSEN AVSGHIQLII PGESACFACA PPLVVAANID EKTTLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPTMSMKPN
301 PQCDNRNCRK QQEYKKKVA ALPKQEVQIE EEEIHHEDNE WGIELVSEVS
351 EEELKNFSGP VPDLPPEGITV AYTIKKQED SVTELTVEDS GESLEDLMK
401 MKNM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c24, frame 1

TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid
 T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1
 YAC yUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus
 fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796_4 gene: "moeB"; product: "MoeB"; Staphylococcus
 carnosus molybdenum cofactor biosynthetic gene cluster, complete
 sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.
 Length = 419

HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122
 Identities = 241/367 (65%), Positives = 293/367 (79%)

```

Query:   37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96
          R +IEK+S+EVVDSNPYSRLMAL+RMGIV++YE+IR  VA+VGVGGSV AEMLTRCG
Sbjct:   48 RQKIEKLSAEVVDSDNPYSRLMALQRMGIVNEYERIREKTVAIVGVGGVGSVVAEMLTRCG 107

Query:   97 IGKLLFDYDKVELANMNRLEFFQHQAGLSKVQAAEHTLRNINPDVLFVHNYNITTVEN 156
          IGKL+LFDYDKVE+ANMNRLEFFQ+QAGLSKV+AA  TL ++NPDV  EVHN+NITT++N
Sbjct:  108 IGKLILFDYDKVEIANMNRLEFFQHQAGLSKVAAARDTLIHVNPVQIEVHNFNITMDN 167

Query:  157 FQHFMDRISNGGLEEGKPVLDVLSVDNFEARMTINTACNELGQTMESGVSENAVSGHI 216
          F F++RI  G L +GK +DLVLSVDNFEARM +N ACNE  Q WMESGVSENAVSGHI
Sbjct:  168 FDTFVNRIKGSITDGK-IDLVLSVDNFEARMAVNMACNEENQIWMESGVSENAVSGHI 226

Query:  217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLFLLNF 276
          Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTT VVAG LV N LK+LLNF
Sbjct:  227 QYIEPGKTACFACVPLVVASGIDERTLKRGVCAASLPTTMAVVAGFLVMNTLKYLLNF 286

Query:  277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDNRNCRKQEEYKKKVAALPKQ-EV-IQEEEEI 334
          G VS Y+GYNA+ DFFP  S+KPNP CDD +C ++Q+EY++KVA  P  EV + EEE +
Sbjct:  287 GEVSQVGVYNALSDFPRDSIKPNPYCDDSHCLQRQKEYEYKVAQPVLDLEVEVPEEETV 346

Query:  335 IHEDNEWGIELVSEVEELKNFSGPVPDLPEGITVAYTIKKQEDSVTELTVEDSGESL 394
          +HEDNEWGIELV+E SE  + S  + G+ AY P K+ D+ TEL+  + +
Sbjct:  347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAYE-PIKR-DAQTELSPAQA--AT 399

Query:  395 EDLMKMKM 403

```


DKF2phtes3_15c6

group: transmembrane protein

DKF2phtes3_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```

1 GAGACACTGA GCCCCGAGAC AGTGAGTGGT GGCCTCACTG CTCTGCCCCG
51 CACCCCTGTCA CCTCCACTTT GCCTTGTTGG AAGTGACCCA GCCCCCTCCC
101 CTTCCATTCT CCCACCTGTT CCCAGGACT CACCCAGGCC CTGCCTGCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGCCTTGGAG CCAGAGACAC
251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGCCTGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCGTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCCTGGCTA GGAGTGTTCG
401 TGTTTCCTCC TTTGCCTACC ACTCTGGGGT GGGGCACTGT GTGGGGAAGC
451 TGGCTGTCTG ATGGTAGCTA TTCCACCCTC TGCCTGCCTG CTGCCTGCT
501 GTCCTGGGCA TGGTGCACTA CCTGTGCCTA GGATTGGTTT TAAATTGTGA
551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGGAGTCCCT
601 TCCCACAGCC TGCCTTGCC TCCCTGCCTC ATCTCTATTC TCATTCCACT
651 ATGCCCAAG CCCTGGTGGT CTGGCCCTTT CTTTTCCTC CTATCCTCAG
701 GGACCTGTGC TGCTCTGCCC TCATGTCCCA CTGGGTGTT TAGTTGAGGC
751 ACTTTATAAT TTTTCTCTTG TCTTGTTTC CTTTCTGCTT TATTTCCCTG
801 CTGTCTCCTG TCCTTAGCAG CTCAACCCCA TCCTTGCCA GCTCCTCCTA
851 TCCCGTGGGC ACTGGCCAAG CTTTAGGGAG GCTCCTGGTC TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGTG GGTGAGCCA GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACCTTC ACTGCATCCT TGCCCATTC AGCCCGGCCT
1001 TTCATGATGC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCCTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGGTGA TCCAAGTGTG GTGGGACCCC CTAAGAGGT CAGGAAGTGG
1151 AACTAATCAT CTGTGCAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGCGGCC GCTCTAGAGG
1251 ATCCAAGCTT ACGTAAAAAA AAAAAAATAA AAG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118
Category: putative protein

```

1 MVAIPPSACL PACCPGHGAV PVPRIGFKFV NNFFFLVDV NRAREVLPTA
51 CACLPAASSLF SFHYAPSPGG LALSFSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCSEFL YFPAVSCP

```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c6, frame 2

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

>PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana
Length = 258

HSPs:

Score = 76 (11.4 bits), Expect = 4.0e-01, P = 3.3e-01
Identities = 30/91 (32%), Positives = 44/91 (48%)

Query: 15 PGHGAVPVPRIGFKFVNNFFGLVDVNRAREVLPTACACLPASSLFSFHYAPSPGGLALS 74
PG GA P+ R+ F+ PF + +E+ A C P SSL+ A G L
Sbjct: 52 PGRGA-PLARVTFRH----FFRF---KKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLP 103
Query: 75 FSSYPQGPVLLCP---HV-PLGCLVEALYNFSLVL 105
S P+G V+ C HV G L A ++++V+
Sbjct: 104 LRSIPEGAVV-CNVEHHVGDGRGLARASGDYAIVI 137

Pedant information for DKFZphtes3_15c6, frame 2

Report for DKFZphtes3_15c6.2

[LENGTH] 118
[MW] 12413.79
[pI] 7.53
[PROSITE] LEUCINE_ZIPPER 1
[PROSITE] MYRISTYL 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] TRANSMEMBRANE 1

SEQ MVAIPPSACLPAACCPGHGAVPVPRIGFKFVNNFFGLVDVNRAREVLPTACACLPASSLF
PRD ccc
MEM
SEQ SFHYAPSPGGLALSFSYYPQGPVLLCPHVPLGCLVEALYNFSLVLCSEFLLYFPAVSCP
PRD eeeeeeeeeeeeeeeeeeeeeeeeecccccccccccccccccccccccccccccccccccc
MEMMMMMMMMMMMMMMMMMMM.

Prosite for DKFZphtes3_15c6.2

PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_15c6.2)

DKFZphtes3_15g14

group: testes derived

DKFZphtes3_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```

1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATTG TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTTAGAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACCT CCAATGTGA AGAAAAAGCT
401 CATGTTTTAA GCTCCTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCTGTG GATGTAAGAG AGAAGTGGCT TTCTAAAACA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAAGAG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTCCATTTT TAGTAACGTG
601 AGGAAAAAAC AGTGAAATTG TTGTAAGAAC AAATCTTGAA TATAAAGAAC
651 TTTGTCAATT GGTATCTGAA GAGGAAGCAT TTGACTTTTT TAAATATTG
701 GATGCAAGA AAGAAAATTC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCCACGA AAGCTGTCC ACCATTTTGT CAACAAAAAG TTTGGAAACC
801 TTGTGGAAAC CAAATCTTTT TCTAAATGA ATTGCAGTGC TGGAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCCGGAA AAAGCACACA AACGTGGGAA
901 AAGGCCCTCT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTACGAAA GGAAACCTG GAAATGTTG AAGCGATTG TTTTTAGCT
1001 ATCAAACCTG GTGTTATTCC TTCGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCTATC AAGCAATGGT TGTTAGAAAA GTGACTCCAG
1101 AGAGGTTGAA AAATATTGAA AAAGAAATG AAAAGAAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATTCCCTG AGACTTGGTC AGCTCAAAGG
1201 AAATCACTTT GATATTGTCA TTAGAAATTT AAAAAACAA ATAAATGATT
1251 CTGCAACCTT GAGGGAGAGA ATTATGGAAG CAATAGAAAA TGTTAAGAAA
1301 AAAGCCTTTG TGAATTACTA TGGACCACAG AGATTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAAATG GACTAGCTTT GCTGAAGAAT GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAAG ACTTGGATGA TCCTGTAAT
1451 AGAGCAAAGA AGTATTTTCT TCAAACTGAG GATGCTAAG GCACACTTTC
1501 ATTGATGCCT GAATTCAAAG TCCGTGAGAG ACCATTGTTG GAGGCATTGC
1551 ACCGCTTTGG CATGACCAGG GAAGGTTGTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTTCCA TGCGCATATT CTATGTTTAC GCATATACCA GCAAAATTTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAA ATTCAGTACC
1851 CGAAGAACAA AGTAGGCGAG TGGTACCATG ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCACCCTTT CCGTTTTTGA AAATTATTGA
2201 TCAGAAACAA ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAAT TTCTCATCCA ATTTTATAT CTTGGCATA GCTGACCTTC
2351 TTGACCATTG GTAATTTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCTA TCTTAGATTG TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTG TGTTGGTACT GATAGAATCA AGACATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTTGAAAA TTTTAATTTT TGTTGGTTTA
2601 ACTATCCAC ATTATAAATC ATCCTTCACC ATTTTATATC AGTTAAATAT
2651 GGGTGTGTTG GGGAGGAATG ACTGGCATGT AGACATGTAT TGATTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAATTTAA

```



```

2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAAAC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTG CGTAACTAAA
2951 GTAAGTAAAA TGATGTGTAT CAACTTGATG GTAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTG GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTCACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAG AGCACAGATA TTGGAGACAA
3151 ACTAACCAG TTTGAACCT GGCCTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAAGTTAT TTAACCTCAT GGGCTTCAGT TTCAACATCT GTAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAT TTTCTAAGTT
3301 AATATATGTA AAGCAACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAA TAAATGAAAA
3401 TCCCTTCTCT TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701
 Category: similarity to unknown protein

```

1 MEEDTDYRIR FSSLCFFNDH VGFHGTIKSS PSDFVIEID EQGQLVNKTI
51 DEPIFKISEI QLEPNNEPKK PKLDLQNLSD EDGRNQEVHT LIKYTDGDQN
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNNE ACDEVREKWL
151 KTELIGLPPE FSIGRILDKN QRASLHSAIR QKFFFLVTVG KNSEIVVKPN
201 LEYKELCHLV SEEEAFDFFK YLDAKKENSK FTFKPDNTKD HRKAVHHEVN
251 KKFGNLVETK SFSKMNCASG NPNVVTVRF REKAHRRGKR PLSECQEGKV
301 IYTAFTLRKE NLEMFEAIGF LAIKLGVIPS DFSYAGLKDK KAITYQAMVV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 KQINDSANLR ERIMEAIEUV KKKGFVNYYG PQRFKGGRKV HTDQIGLALL
451 KNEEMKAIKL FLTPEDLDDP VNRAKKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHRRFGM TEEGCIQAWF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDIDEN FPNSKIHLVT EEEGSANMYA IHQVVLPLVG
601 YNIQYPKNKV GQWYHDILSR DGLQTCREKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTGSHI DETALSLIS FDLASCYAT VCLKEIMKHD
701 V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15g14, frame 2

TREMBL:SPBC1A45P_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid clA4 left hand region 1-26184 bp
 Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)
 Length = 676

HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
 Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSEIVVKPNLEYKELCHLVSEEEAFDFFK-YLDAKKENSKFTFKPDNTKDHHRKAVHHEV 249
 + E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPCLR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNLVETKSFskMNCsAGNPVNVVTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307
+ F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNRTNKQEKINQTRDANGVENWGYGPKDFIHFTL 224

Query: 308 RKENLEMFEAIGFLAIKLGVIPSD-FSYAGLKDKKAITQAMVVVRKVTPERLKNIEKEIE 366
KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVSSISKIGLDRNLNRL- 282

Query: 367 KKRNVNFIKRSVDDSLRLGQKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKKGfV 426
K M + N D SL LG LKGN F +VIR++ N +L E + + + + GF+

Sbjct: 283 -KGMIIIGNYNFSDASLNLGDLKGNFVVVIRDVTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGQPRFGKGRKVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNR-AKKYFLQTEDAK 485
NY+G QRFG + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFGTG-SISTHTIGRELLLSNWKKAELILSDQDNVLPKSKEARKIWAETKDA 399

Query: 486 GTLSLMPEFKVRERALLEALHRFGMTEEGCIQ--AWFS----LPHSMRIFYVHAYTSKI 539
L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPROCLAENALLYSLSNQRKEEDGTYSNAYYTAIMKIPRNLRTMYVHAYQSYV 459

Query: 540 NEAVSYRLETYGARVVQGDVLC-----LDEDIDDENFPNS-----KIHLVTEEGS 585
N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NSIASKRIELHGLKLVVGDVIDTSEKSPISIGDDEDFEDVREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVVLPLVLYNIQYPPKN-VGQWYHDILSRDGLQTCRFKVPVTLKLNIPGCYRQI 644
+ Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLSPGFDVLYPSNEELKQLYVDILKADNMDFNMRRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671
++ P +L Y+++ D + + +D

Sbjct: 580 IQPKSLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01
Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFIVIEIDQGLVNKTIDEPIFKISEIQLEPNFPPKPKLDLQNLSE 81
GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIKRYTDGDNHQSGS--EKEDTI-VDGTSKCEEKADVLSSFLDEKTHELLN 138
R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV-----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKI 161

Query: 139 NFACDVREKWLKTELIGLPPE-FSIGRILDKNQASLHSAIRQ 181
+RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698
++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKFLGTSAYATMALRELMK 660

Pedant information for DKFZphtes3_15g14, frame 2

Report for DKFZphtes3_15g14.2

[LENGTH] 701
[MW] 80700.96
[pI] 7.31
[HOMOL] PIR:S67136 hypothetical protein YOR243c - yeast (*Saccharomyces cerevisiae*) 2e-51

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53
[BLOCKS] BL01268C
[BLOCKS] BL01268B
[BLOCKS] BL01268A
[SUPFAM] hypothetical protein HI0701 3e-06
[PROSITE] MYRISTYL 7
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 16
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 13
[PROSITE] ASN_GLYCOSYLATION 5
[KW] Alpha_Beta


```

SEQ  MEEDTDYRIRFSSLCFFNDHVGFGHTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEI
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QLEPNFPPKPKLDLQNLSELDGRNQEVHTLIKYTGDQNHQSGSEKEDTIVDGTSKCEE
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  KADVLSSFLDEKTHELLNNFACDVREKWLSKTELIGLPPEFSIGRILDKNQRASLHSAIR
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEEAFDFFKYLDAKKENSKFTEKPDNTKD
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  HRKAVHHFVNKKFGNLVETKSFSKMNCAGNPVNVTVRFREKAHKGKRPPLSECQEGKV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  IYTAFTLRKENLEMFEAIGFLAIGLVIPSDFSYAGLKOKKAITYQAMVVRKVTPERLKN
PRD  eeeeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  IEKEIEKKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KKKGFVNYYPQRFQKGRKVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNRAKKYFLQ
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TEDAKGTLSLMPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRIFYVHAYTSKIWN
PRD  hcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  EAVSYRLETYGARVVQGDVLCLEDDIDENFPNSKIHLVTEEGSANMYAIHQVVLPLVLG
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  YNIQYPKNKGQWYHDILSRDGLQTCRFKVPKTLKLNIPGQYQILKHPNLSYQLMEDHD
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  IDVKTKGSHIDETALLISFDLDASCYATVCLKEIMKHDV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15g14.2)

DKFZphtes3_15h1

group: testes derived

DKFZphtes3_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACCAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
51 TCTCTAAATC CGGTCACCAT GTCGGACCCC GAAGGCGAGA CCTTGCGAAG
101 CACCTTTCCC TCTTATATGG CCGAAGGCCG GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCCTGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTTCTG TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCCAG GAAGCCATCA
451 ACAACTCAGT GGGAAAGTCCT TCTTCCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAAT ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATG AAACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGGA
601 AGGCTCTCGT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGGAGCTC
651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGGCG GCCTGACTGT GGAGGACCTC ATCATGACGG
751 GCATCAACTA CCTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTGCG
851 GGACCACAAA CGCCGTCCCT CACAGACAGC CCATTACATC CTCAGAGGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGTCTT
951 CAGAAAGCTG AGAAAGTGCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAAC AAGGATGAAC TGGTTGGAAG CTTGTATAGC TGCATAGGGA
1051 ATGCCCATG TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTGTGC CAGAGTTGGG AAATTCCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAAGATCC CTCTGGCAAA AACCACCCTG
1251 GAGAAGACCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCCTGG CAGGCCCAGA ATTATGGCGA GAAGTCCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGCACAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGGTATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAAA GAAAAAAGCG AGGGAGAAGC
1601 TTCCTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCGAGAAG GTGGTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAAGGGC CTGTCAGGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CTTCCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGCAGAGG ACCAGAAGAA CTAAAGAAAC TTTCAGAAAT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAAATT GGAGAAATAG GAGAAACGAA
2051 AAAACAGGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTATTATAA
2151 CTGGATTTTC AAGCGATTG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACTCTT
2251 GCAAAAAAAA AAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672
 Category: similarity to known protein

```

1 MSDPEGETLR STFPSYMAEG ERLYLCGEFS KAAQSFSNAL YLQDGDKNCL
51 VARSKCFLKM GDLESLKDA EASLQSDPAF CKGILQKAET LYTMGDFEFA
101 LVFYHGRGYKL RPDREFRVGI QKAQEAENNS VGSPSSIKLE NKGDLSEFLSK
151 QAENIKAQOK PQPMKHLHP TKGEPKWKAS LKSEKTVRQL LGELYVDKEY
201 LEKLLDDEDL IKGTMKGGLT VEDLIMTGIN YLDTHSNFWR QOKPIYARER
251 DRKLMQEKWL RDHKRRPSQT AHYILKSLED IDMLLTSGSA EGSLQKAEKV
301 LKKVLEWNKE EVPNKDELVG NLYSCIGNAQ IELQMEAL QSHRKDLEIA
351 KEYDLPDAKS RALDNIGRVF ARVGKFOQAI DTWEEKIPLA KTTLEKTWLF
401 HEIGRCYLEL DQAWQAQNYG EKSQCCAESE GDIEWQLNAS VLVAQAQVKL
451 RDFSASVNNF EKALERAKLV HNNEAQAII SALDDANKGI IRELKRTNYV
501 ENLKEKSEGE ASLYEDRIIT REKDMRRVRD EPEKVVQWQD HSEDEKETDE
551 DDEAFGEALQ SPASGKQSV EAGKARSDLGA VAKGLSGELG TRSGETGRKL
601 LEAGRRESRE IYRRPSGELE QRLSGEFSRQ EPEELKKLSE VGRREPEELG
651 KTQFGEIGET KKTGNEMEKE YE

```

BLASTP hits

Entry AF039202_1 from database TREMBL:
 product: "Hsp70/Hsp90 organizing protein"; *Cricetulus griseus*
 Hsp70/Hsp90 organizing protein mRNA, complete cds.
 Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782_1 from database TREMBL:
 product: "myosin heavy chain"; *Argopecten irradians* myosin heavy chain
 mRNA, complete cds.
 Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:
 stress-induced protein stil - soybean
 Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3_15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15h1, frame 3

Report for DKFZphtes3_15h1.3

```

[LENGTH]      672
[MW]           76655.61
[pI]           5.49
[HOMOL]        PIR:S56658 stress-induced protein stil - soybean 6e-10
[SUPFAM]       tetratricopeptide repeat homology 1e-07
[PROSITE]      MYRISTYL 7
[PROSITE]      AMIDATION 3
[PROSITE]      CAMP_PHOSPHO_SITE 4
[PROSITE]      CK2_PHOSPHO_SITE 15
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 11
[PROSITE]      ASN_GLYCOSYLATION 2
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 4.76 %

```

```

SEQ  MSDPEGETLRSTFPSYMAEGERLYLCGEFSKAAQSFSNALYLQDGDKNCLVARSKCFLKM
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  GDLESLKDAEASLQSDPAFCGILQKAETLYTMGDFEFALVIFYHGRGYKL RPDREFRVGI
SEG  .....
PRD  hcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```


Prosites for DKFZphtes3_15h1.3

593

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_15h1.3)

DKFZphtes3_15i5

group: cell structure and motility

DKFZphtes3_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results: No predictive prosite, pfam or SCOP motife.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)
"radial spokehead" part of flagella in *Chlamydomona*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```

1  CACCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAACGGCC CCCTCTCTCG
51 GTGCTCAGAA ACCGGCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101 TCTGCAAGCC TTCTCTCTAG AGATCTGTGC CTCTTGCGGA ACCATGGGAG
151 ACCTGCGCGC CTACCTGAG CGCCCTGCCC AGCAGCCTCC GGGCCGGAGG
201 ACTTCTCAGG CTTCCACAG GCGGCACAGT CGGGACCAAG CTCAGGCCCT
251 GGCAGCGGAG CCGGAGGAGA GGCAGCAGAT ACCTCCAGAC GCCCAGCGAA
301 ACGCCCTTGG TTGGTCACAG AGGGGCAGCC TGTCCCAACA GGAGAACTTG
351 CTGATGCCCC AGGTCTTCCA GGCTGAGGAA GCCCGGCTGG GTGGCATGGA
401 GTACCCATCT GTGAACACGG GCTTTCCTTC AGAGTTCCAG CCTCAGCCTT
451 ACTCTGATGA AAGCAGGATG CAGGTCGCCG AGCTCACCAC CAGCCTAATG
501 CTGCAGCGGC TCAGCAGGGG CCAAGCAGC CTGTTCAGC AACTGGACCC
551 CACCTTCCAG GAGCCCCCAG TCAACCCCTT GGGCCAGTTC AACCTCTACC
601 AGACAGACCA GTTCTCTGAA GGTGCCCAGC ACGGGCCTTA CATAAGGGAT
651 GACCCCTGCC TTCAGTTCTT GCCCTCTGAG CTGGGCTTCC CACACTACAG
701 TGCCCAAGTG CCTGAGCCCG AGCCTCTGGA GCTGGCCGTG CAGAACGCCA
751 AGGCCTACCT GCTGCAGACC AGCATCAATT GCGACCTCAG CCTGTACGAG
801 CACCTGGTAA ATCTGTCTGAC CAAGATCCTG AACCAGCGGC CTGAGGACCC
851 CTTGTCTGTC CTGGAGTCTC TGAACCGCAC CACGCAGTGG GAGTGGTTCC
901 ACCCCAAGCT GGACACGCTG CGGGACGACC CCGAGATGCA GCCCACCTAC
951 AAGATGGCGG AGAAACAGAA GGGGCTGTTC ACCCGGAGTG GAGGCGGCAC
1001 TGAAGGCGAA CAGGAGATGG AGGAGGAGGT GGGGGAGACA CCAGTGCCCA
1051 ACATCATGGA GACTGCCTTC TACTTCGAGC AGGCCGCGCT CGGCCTGAGC
1101 TCGGACGAGA GCTTCCGCAT TTTCCTGGCC ATGAAACAGC TGGTGCAGCA
1151 GCAGCCCATC CACACCTGTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACCT GGTGGCCGAG GTGGAATTCC GGGAGGGCGA GGAGGAGGCA
1251 GAGGAGGAGG AGGTGGAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCTTAAGTC CGTATGGAAG CCGCCGCCCG TGATCCCCAA GGAGGAGAGC
1401 CGCTCAGGCG CCAACAAGTA CCTGTACTTT GTGTGCAACG AGCCGGGCCT
1451 GCCATGGACG CGGCTGCCCC ACGTCACTCC AGCCCAGATC GTGAACGCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACACGCC AGTCGTCAGC
1551 TACCCACCCT TCCCGGGCAA CGAGGCCAAC TACCTGCGGG CCCAGATAGC
1601 CCGCATCTCG GCCGCCACGC AGGTCAGCCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAAG GTGGTGCTGG GCGCGACTCC
1701 TACGAGGAGA ACCCGGACTT CGAGGGCATC CCCGTGCTGG AGCTGGTCTG
1751 CTCCATGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGCAGGGCC
1801 GCTGCACCTG GGTGAACCCCT TTGCAGAAGA CAGAGGAGGA GGAGGACCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG TGGAGCAGGA
1901 GGTTGGCCCC CCACTGCTAA CGCCACTTTC AGAAGATGCA GAAATCATGC
1951 ACCTGGCACC CTGGACCACC CGCCTGTCTT GCAGCCTCTG CCCGCAGTAC
2001 TCACTGGCCG TTGTGCGCTC CAACCTCTGG CCCGGGGCCT ATGCCATATG
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTACA
2101 GCCCGGAGAG CTTCAACCCG GCCCTGCCAG CCCCATTCA ACAAGAGTAC
2151 CCCAGTGGCC CAGAGATCAT GGAGATGAGT GACCCACAGC TGGAGAGGGA
2201 GCAGGCTCTG AAAGCAGCCC AGGAACAAGC CCTGGGAGCC ACAGAGGAGG
2251 AGGAGGAGGG CGAGGAGGAG GAGGAGGGCG AGGAGACAGA TGAAGAGGC

```


2301 CCACCCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCTCTTA
 2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCCAGA
 2401 GGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAATAAAA TTCCTCCACG
 2451 GCATTAAAAA AAAAAAAAAA AAAAAAAG

BLAST Results

No BLAST result

Medline entries

86251010:
 Molecular cloning and expression of flagellar radial spoke and dynein genes of *Chlamydomona*

81142496:
 Radial spokes of *Chlamydomonas* flagella: polypeptide composition and phosphorylation of stalk components.

9450971:
 Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm axonemes: involvement of the protein in the regulation of sperm motility.

Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717
 Category: strong similarity to known protein

1 MGDLPYPYPER PAQPPGRRT SQASQRRHSR DQAQALAADP EERQQIPDDA
 51 QRNAPGWSQR GSLSQENLL MPQVFQAEAA RLGGMEYPSV NTGFPSEFQP
 101 QPYSDESRLQ VAEITTSMLL QRLQQGSSSL FQQLDPTFQE PPVNPLGQFN
 151 LYOTDQFSEG AQHGPYIRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ
 201 NAKAYLLQTS INCDLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTTQWE
 251 WFHPKLDTLR DDPQMPTYK MAEKQKALFT RSGGGTEGEQ EMEEEVGETP
 301 VFNIMETAFY FEQAGVGLSS DESFRIFLAM KQLVEQQPIH TCRFWGKILG
 351 IKRSYLVAEV EFREGEEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV
 401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV
 451 NARKIKKFFT GYLDPVVS Y PFPNGEANY LRAQIARISA ATQVSPGLGFY
 501 QFSEEEGDDE EGGAGRDSY EENPDFEGIP VLELVDSMAN WVHHTQHILP
 551 QGRCTWVNPL QKTEEEEDLG EEEKADDEGP EEVEQEVGPP LLTPLSEDAE
 601 IMHLAPWTTR LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH
 651 KYSPESFNPA LPAPIQQEYP SGPEIMEMSD PTVEEEQALK AAQEALGAT
 701 EEEEEEEEE EGEETDD

BLASTP hits

Entry U73123_1 from database TREMBL:
 product: "radial spokehead"; *Strongylocentrotus purpuratus* radial spokehead mRNA, complete cds.
 Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:
 radial spoke protein 6 - *Chlamydomonas reinhardtii*
 Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15i5, frame 3

Report for DKFZphtes3_15i5.3

{LENGTH} 717
 {MW} 80913.61
 {pI} 4.36

[HOMOL] TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus
radial spokehead mRNA, complete cds. 1e-130
[PROSITE] TRANSFERRIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 21.48 %

```

SEQ  MGDLPYPYPERPAQPPGRRTSQASQRRHSRDQAQALADPEERQQIPPDQAQRNAPGWSQR
SEG  .....XXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccc

SEQ  GSLSQENLLMPQVFQAEEARLGM EYPSVNTGFPSEFQPPYSDSRMQVAELTSLML
SEG  .....XXXX
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccc

SEQ  QRLQGGQSSLFQQLDPTFQEPVNPVGQFNLYQTDQFSEGAQHGPYIRDDPALQFLPSEL
SEG  .....XXXXXXXXXX.....
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPELELAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNQRPEDPLSVL
SEG  .....
PRD  cccccccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhccccchhhh

SEQ  ESLNRTTQWEWFHFKLDTLRDDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEVGETP
SEG  .....XXXXXXXXXXXXXXXXXXXXX...
PRD  hhhchhhhhccccccccccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhcccc

SEQ  VFNIMETAFYFEQAGVGLSSDESFRIFLAMKQLVEQQPIHTCRFWGKILGIKRSYLVAEV
SEG  .....XXX
PRD  cccccccccccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhccccchhhhhhhhh

SEQ  EFREGEEEAEVEEEMTEGGEVMEAHGEEGEDEEKAVDIVPKSVWKPVPVIPKEESR
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhhhhhccccccccccccccccchhhhhheeecccccccccccccccccccc

SEQ  SGANKYLYFVCNEPGLPWTRLPHVTPAQIVNARKIKKFTGYLDTPVVSYPFPFGNEANY
SEG  .....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccccccchhhh

SEQ  LRAQIARISAAQTQVSPGLGFYQFSEEGDEEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhhh

SEQ  WVHHTQHILPQGRCTWVNPLOKTEEEEDLGEEEKADEGPVEVEQEVGPPLTLPLSEDAE
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  hhhccccccccccccccccchhhhhhhhhccccchhhhhcccccccccccccccccccccccccc

SEQ  IMHLAPWTTRLSCSLCPQYSVAVVRNLWPGAYAYASGKKFENIYIGWGHKYSPEFNPFA
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LPAPIQEQEYPSGPEIMEMSDPTVEEQALKAAQEQALGATEEEEEEGBEEEGEETDD
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

```

Prosite for DKFZphtes3_15i5.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKFZphtes3_15i5.3)

DKFZphtes3_15j18

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGGCGGCCCG TCTGCCGGCC TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATTT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACTCAGGT GGAATGCCCT ACCCCAGTCC TGCGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACCCC CTCATCCCTG CACCCCTTCC CTGCCTGGGA
451 TTCTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTTCCTAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGTCTCTCTT TGCTTCCAG CCAGTGTTTT GGGGTTTGAA
601 GTTGGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTCTCAT
751 CAGGTGATTC TTGGTTGTAG CAAAGTTCAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTTG TGAAACTCTA AAAAAAAAAA
851 AAAAAAAG GGCAGCGCGT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2
-----ORF from 110 bp to 553 bp; peptide length: 148
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPAWRD VLQPGVGGE GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHPFSEGV GTQVECLTPV LRLEDMART
101 APHPSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

BLASTP hits

No BLASTP hits available.

Alert BLASTP hits for DKFZphtes3_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15j18, frame 2

Report for DKFZphtes3_15j18.2

[LENGTH] 148
 {MW} 15665.78
 {pI} 8.91
 {PROSITE} MYRISTYL 3
 {PROSITE} CK2_PHOSPHO_SITE 1
 {KW} Irregular

SEQ MFGCPVRCFKPPTQLISGEASARLPAPWRDVLQQPGVGEGGLRISWQGAPKSRVRPAFI
 PRD cccccccccccccccccccccccccchhhhhhcccccccccccccccccccccccccccc

SEQ SPVPFTVLQSQHYHPFSEGVGTQVECLTPVLRLESDMARTAPHPSSLHPFPAPWDSSSPVH
 PRD cccccccccccccccccccccccccchhhhhhcccccccccccccccccccccccccccc

SEQ CGAPLPSAHGGFPRARAEGSWSQPGAGS
 PRD ccccccccccccccccccccccccccccccccccccc

Prosites for DKFZphtes3_15j18.2

PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	49->55	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15j18.2)

DKFZphtes3_15j3

group: nucleic acid management

DKFZphtes3_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276c, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits
YGR276c = ribonuclease H
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp
Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CCTTCTTGCG CAGGCAGACG
151 CCCGTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGGAAAGGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAAG CAGGCAGGCC
251 CCAAAATAAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTAAGTACAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
401 TATGCAGTTC TGGGCAATC CAATGTTCCT AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACCACTAA ACAACGTAGT GGTTTTGTT CTGCAGGGAA
501 TGAGTCAGCT ACACTTTAC AGGTTCTATT TGGAGTTTGG ATGCTTTCGA
551 AAAGCATTCG GACATAAATT CCGCTTGCCCT CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAATGAACA AAGAGCTGGA GATCTGCCCA
651 AGACAATGGA AGGGCCTTTA CCTTCTAATG CAAAGCCGCG CATCAACCTT
701 CAGGATGATC CCATCATTC AAGATATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCGCT CTGACAAAGG AGGAAATGAG AACGTTTCAC TTTCCATTAC
801 AAGGTTTTCC TGATTGTGAA AACTTTTAC TTACCAATG TAATGTTTCT
851 ATAGCAGACA ATAGTCCTCT CTTTGGACTT GACTGTGAAA TGTGCTCAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGTTGTCT GAAGGAGGCT
951 GCTGTGTTAT GGATGAACTG GTCAAACCTG AAAACAAGAT TCTGGACTAC
1001 CTCACCAAGT TTTCCGGAAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTGAG ATCTCAGAGC ACTGAAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAGT TCTTAGCCAA AGTTATTTTG GGAAGGATA
1251 TACAGTGTCG AGACAGACTT GGTCTATGAT CCACAGAAGA TGCTAGAAC
1301 ATCCTTGAAT TGGCTCGGTA TTTCTTAAAG CATGGCCCAA AAAAGATTGC
1351 AGAACTAAAT CTAGAAGCAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCAGCC AAACACAAGT
1451 GTTTTGAAT GCTTGGATTG AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGCTCTTC AAATAAGAGG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGGCCT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAAGTGTCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTTAAAGCTT TGGCCAGTGC CAGTCAATGA CTTTGTCTCT
1801 TGAACCCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCAGC CTGATTGTG
1851 ACACCCCTCGT GAATGAGCTG GAAGGAGATT CTGAAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCTCTCTT CTTGGCCTGG AAGCTGTGAT CTGCTCTAAA GATCTTAAAA
2001 GTGGAAAGCA GAAAAATAC TGTTCCTGA AATTCAAAG TTTTGGCAGT
2051 GCCCAGCAGG CCTCAACAT TCTCACAGGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCCA GGCACCTCCA TGCCCTGGCT AGAGGCTTAC
2151 CACCTGAATC AACAAAGGCT CCAGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTGCAGAC TCTGAACTG GACCACCGA AGATAGCAGC
2251 CTGGCGCTGG AGCCGGAAGA TTGGAAGACT CTACAACAGC TTGTGCCCGG
2301 GCACCTCTCG CCTCATCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGAATAAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCCGT TGTTCGTGAG TCGGCTGCC ATGTTTCCAT GTGCCATTC
```


2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTCAGGCTG TAGCCTCCCC
 2501 AACACGAGA CAGTTTTATG GAACTTGGT ATAGCAGCTA AAAGAGTTTA
 2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAG
 2601 CATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
 2651 AGGCGCGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743
 Category: similarity to known protein

1 MEPEREGTER HPRKVRRESRQ APNKLVGAAE AMKAGWDLEE SQPEAKKARL
 51 STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
 101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
 151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVGLTRC LLTKEEMRTF
 201 HFPLQGFPCD ENFLTKCNG SIADNSPLFG LDCMCLTSK GRELTRISLV
 251 AEGGCCVMDE LVKPENKILD YLTSFSGITK KILNPVTTKL KDVQRQLKAL
 301 LPPDAVLVGH SLDLRLALK MIHPYVIDTS LLYVREQGRR FKLKFLAKVI
 351 LGKDIQCPDR LGHDATEDAR TILELARYFL KHGPKKIAEL NLEALANHQE
 401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
 451 CQTKKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRK
 501 WTEISTVYAG PFSKNCNLRA LKRLFKSEGP VQSMTFVLET RQVQRPVTEL
 551 TLDCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
 601 KDLKSGQKQK YCFLKFKSFG SAQALNILT GKDWKLKGRH ALTPRHLHAW
 651 LRGLPPESTR LPGLRVVPPP FEQELQTLK LDHPKIAAWR WSRKIGKLYN
 701 SLCPGTLCLI LLPGKSTHG SLGLGLMGI KEEESAGPG LCS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15j3, frame 2

TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product";
 Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,
 N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid
 C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637_9 gene: "SPAC637.09"; product: "putative
 exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P
 = 2.8e-27

>TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo
 sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.
 Length = 547

HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 358/373 (95%), Positives = 358/373 (95%)

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 164
 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN
 Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 60
 Query: 165 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCDENFLTKCNGSIAD 224
 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCDENFLTKCNGSIAD
 Sbjct: 61 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCDENFLTKCNGSIAD 120

Query: 225 NSPLFGLDCM-----CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 269
 NSPLFGLDCM CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL
 Sbjct: 121 NSPLFGLDCMARTTFNFSIGVLAECCLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 180

Query: 270 DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT 329
 DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT
 Sbjct: 181 DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT 240

Query: 330 SLLYVREQRRFKLFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE 389
 SLLYVREQRRFKLFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE
 Sbjct: 241 SLLYVREQRRFKLFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE 300

Query: 390 LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR 449
 LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR
 Sbjct: 301 LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR 360

Query: 450 NCQTIKCLSNKEV 462
 NCQTIKCLSNKEV
 Sbjct: 361 NCQTIKCLSNKEV 373

Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 175/179 (97%), Positives = 177/179 (98%)

Query: 538 LETRQVQRPVTELTLDCTLVNELEGDSENQGSYLSGVSETFKEQLLQEPRLFGLGAV 597
 L ++VQRPVTELTLDCTLVNELEGDSENQGSYLSGVSETFKEQLLQEPRLFGLGAV
 Sbjct: 368 LSNKEVQRPVTELTLDCTLVNELEGDSENQGSYLSGVSETFKEQLLQEPRLFGLGAV 427

Query: 598 ILPKDLKSGKQKCYFLKFKSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE 657
 ILPKDLKSGKQKCYFLKFKSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE
 Sbjct: 428 ILPKDLKSGKQKCYFLKFKSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE 487

Query: 658 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRSRKIGKLYNSLCPGTLCLILLPGTK 716
 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRSRKIGKLYNSLCPGTLCLILLPGTK
 Sbjct: 488 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRSRKIGKLYNSLCPGTLCLILLPGTK 546

Pedant information for DKFZphtes3_15j3, frame 2

Report for DKFZphtes3_15j3.2

[LENGTH] 743
 [MW] 83536.58
 [pI] 8.87
 [HOMOL] TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
 Chromosome 16 BAC clone CIT987SK-44M2, complete sequence. 0.0
 [FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YGL094c] 1e-10
 [FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YGL094c] 1e-10
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 1
 [PROSITE] CK2_PHOSPHO_SITE 8
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 16
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] RNA_recognition_motif. (aka RRM, RBD, or RNP domain)
 [KW] Alpha_Beta

SEQ MEPEREGTERHPRKVRRESQAPNKLVGAAEAMKAGWDLEESQPEAKKARLSTILFTDNCE
 PRD ccchhhhhccccchhhhhhhcchhhhhhhhhccccccccccccchhhhhhhccccccccce

SEQ VTHDQLCELLKYAVLGKSNVPKPSWCQLFHQNHNNVVVFLQMSQLHFYRFLYFEGCL
 PRD eehhhhhhhhhhhhhhhccccccccceeeccccccccceeeccccchhhhhhhhhhhhhhh

SEQ RKAFRHKFRLPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSNAKAAILQDDPIIQKY
 PRD hhhhhhhhhccccccccchhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhcccccccc

SEQ GSKKVGLTRCLLTKEEMRTFHFPLQGFPCDENFLTKNGSIADNSPLFGLDCMCLTSK
 PRD cccccchhhhhhhhhhhhhhhcc

SEQ GRELTRISLVAEGGCCVMDLVKPKENKILDYLTFSFGITKKILNPVTTKLKDVQRQLKAL
 PRD cchhhhhheeeccchhhhhhhhhhhhh

Prosites for DKF2phtes3_15j3.2

Pfam for DKFZphtes3 15j3.2

604

DKFZphtes3_15k11

group: signal transduction

DKFZphtes3 15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCCGCGCTG CCAACCCTCC
51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGGT CCGGGTGGGG
151 TTCTACGACA TCGAGGGCAC GCTGGGCAAG GGCAACTTCG CTGTGGTGAA
201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAATCTA CCGAGAAGTA
301 CAAATAATGA AAATGTTAGA CCACCCTCAC ATAATCAAAC TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATTCTGGCA AATCCTGTCT GCTGTTGATT ATTGTCTATG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAATCTC CTGCTGGATA
551 ACAACATGAA TATCAAATA GCAGATTTCG GTTTTGGAAA TTTCTTTAAA
601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCTT ATGCAGCCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAGG ACCACAGCTG GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCCTTGTCT GTGGAGCTCT GCCCTTTGAT
751 GCACCGACTC TTCCAATTTT GAGGCAGAGG GTTCTGGAAG GAAGATTCCG
801 GATTCCGTAT TTCATGTCAG AAGATTGCCA GCACCTTATC CGAAGGATGT
851 TGGTCTTAGA CCCATCCAAA CGGCTAACCA TAGCCCCAAT CAAGGAGCAT
901 AAATGGATGC TCATAGAAGT TCCTGTCCAG AGACCTGTTT TCTATCCACA
951 AGAGCAAGAA AATGAGCCAT CCATCGGGGA GTTTAATGAG CAGGTTCTGC
1001 GACTGATGCA CAGCCTTGGA ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1051 CAGAACAAGA GCTATAACCA CTTTGCTGCC ATTTATTCTT TGTGGTGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTGATG
1151 GCGCGCAGGG TCGGCCTAGC ACCATTGCTG AGCAAAAGT TGCCAAGGCA
1201 CAGACTGTGG GGCTCCAGT GACCATGCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCCTCCCCC AGGCATCCAA CGTGGAGGCC TTTTCATTTC
1301 CAGCATCTGG CTGTCAAGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCCCTGTG CTCCTGTCTT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCCAG CAACATGATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCGCTCAT
1501 GCCTTTGAGG CATTTTCAGT CACACGCAGC GGGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCGTGAT GCCTGGGGCA GGGAAAATTT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGAGTATGAT
1651 ATGGGGTCTG TTCAGAGGGA CCTGAACTTT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGCATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCAGCCA TGCAGGCTCT GAGCTCCCAG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGGCCCGCAG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATT AGACAACATC
1901 TTCAGAATCT GGCTAGAACC AAAGGAATTC TAGAGTTGAA CAAAGTGCAG
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCCTAACC TGGCGCCGGC
2001 GGCTCCTCAG CTCCAGGACC TTGCTAGCAG CTGCCCTCAG GAAGAAGTTT
2051 CTCAGCAGCA GGAAAGCGTC TCCACTCTCC CTGCCAGCGT GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCCAGC CTCTGTGCAA AGGCCAGAA CACCTGTGAG CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CACTGCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCTTCCCCCG CCAGGAGACT CCACCGCCTT CTCAGCAGGC CCCACCGTTC
2401 AGCCTGACCC AGCCCTGAG CCCCCTCTG GAGCCTTCTT CCGAGCAGAT
2451 GCAATACAGC CCTTCTCTCA GCCAGTACCA AGAGATGCAG CTTTACGCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGCTC CTCCTCTGCC CACGCAGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCACCC CCTCCACCAC CACGACAGCC
2601 AGGAGCTGCC CCAGCCCCCT TACAGTTCTC CTATCAGACT TGTGAGCTGC
```



```

2651 CAAGCGCTGC TTCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAGAAG
2751 CCCAGGACTG CAAGAGGCCCT CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACGCTGT GGATCCACAA
2851 CACAACGGGT ATGTCCTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTTCTTAT TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGGT GGATGTTGCT TCCTCCTGGT
3051 TCTGCCCCAC CACAAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGGTC CTGCCCTTCG GTCGAGTGGA GCAAGCTCTC GAGGGCAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATTATCAA GGGCAACCTT GGTGAAAGCA GAAAGGGTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCACTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCAATTG AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGGCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCAG GGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCCTGCACT TTCTCGTGGG CACTGCTGGC TTGCGGGCAG TCGGTCTCCA
3751 GGGTACCTGT TGTCTCTTTT CCGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTGGAC AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAGG ACATGATAAA TGAACTTTTC TGTCCCATGT
3901 GAAGTGGTAG TGCGGTGCCT TTCCCCCAGA TCATGCTTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTTATG TCAATGTCAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTTCCT CCATGTGGGA ATCAGCATTC TTAATTTCTG
4051 TAAAGTTTTG ACTTGTAAAT AAATGTTCAA GTATTACAGC AATATTCAAA
4101 GAAAGAACCA CAGATGTGTT AACCATTATA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTA ACCAACACTT ACAATTCACT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTTGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGTG
4301 AATTCTAAAG ATGATCATT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAAACACCT TTAAGTTGCC TTTCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGGTTCCATT
4501 CTGGCCATGC GAGCCAGCT CCTACCAACG TCGGTAACCT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTCGCCA GCGCTCACC TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACTG TGCTCAGCGT TGGACTTGGT GTCAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTGTG ATTAAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAAATA AAGAACATCA TTTCAATTAA
4801 AAAAAAAAAA AAAAAAAGG CGGCCGCTCT AGAGGATCCA AGTTACGTA
4851 CGCGTAAAAA AAAAAAAG

```

BLAST Results

Entry HSG4921 from database EMBL:
human STS SHGC-37164.
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:
Homo sapiens mRNA for KIAA0781 protein, partial cds.
Score = 10725, P = 0.0e+00, identities = 2145/2145

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959
Category: known protein

```

1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQGPVVRVG
51 FYDIEGTLGK GNFAVVKLGR HRITKTEVAI KIIDKSQDLA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVMEY KSMYLVTEY AKNGEIFDYL ANHGRLENESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPFYAAPEVF EGQYEGPQL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSDCEHLI RRMLVLDPK RLTIQIKHEH

```



```

301 KWMLEIEVPVQ RPVLVPQEQE NEPSIGEFNE QVLRMLHSLG IDQOKTIESL
351 QNKSYNHFAA IYFLLVERLK SHRSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTMH SPNMRLRSA LLPQASNVEA FSFPASGCQA EAAFMEEECV
451 DTPKVNGLL DPVPPVLVRK GCQSLPSNMM ETSIDEGLET EGEAEEDPAH
501 AFEAFQSTRS QRRHTLSEV TNQLVVMPGA GKIFSMNDSP SLDSDVSEYD
551 MGSVQRDLNF LEDNPSLKD I MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVNHRSPV SFREGRRASD TSLTQGIVAF RQHLQNLART KGILELNKVQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPAVHPQ
701 LSPRQSLETQ YLQHRLLQKPS LLSKAQNTCQ LYCKEPPPSL EQQLQEHRLQ
751 QKRLFLQKQS QLQAYFNQMQ IAESSYPQPS QQLPLPRQET PPSQQAAPPF
801 SLTOPLSPVL EPSSEQMYS PFLSQYQEMQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPRQPGAA PAPLQFSYQT CELPSAASPA PDYPTPCQYP
901 VDGAAQSDLT GPCPRSPGL QEAPSSYDPL ALSELPLGLFD CEMLDVADPQ
951 HNGYVLVN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15k11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15k11, frame 1

Report for DKFZphtes3_15k11.1

```

[LENGTH] 926
[MW] 103915.77
[pI] 5.70
[HOMOL] TREMBL:AB018324.1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```


[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]
2e-14

[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14

[FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12

[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YBR097w] 1e-10

[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]
1e-10

[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]
1e-10

[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]
4e-09

[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.
cerevisiae, YHR079c] 1e-07

[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c]
1e-07

[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04

[BLOCKS] BL00415A Synapsins proteins

[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins

[BLOCKS] BL00107A Protein kinases ATP-binding region proteins

[SCOP] dlgol_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 3e-78

[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens) 1e-81

[SCOP] dikoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis 5e-89

[SCOP] dikoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har 5e-86

[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 3e-80

[SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 6e-70

[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 1e-95

[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 7e-71

[SCOP] dlydse_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 2e-96

[SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 2e-72

[SCOP] dlcdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 5e-97

[SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma 2e-68

[SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 3e-53

[SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 3e-78

[SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 1e-58

[EC] 2.7.1.117 Myosin-light-chain kinase 3e-49

[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 4e-78

[EC] 2.7.1.38 Phosphorylase kinase 3e-41

[EC] 2.7.1.37 Protein kinase 7e-45

[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42

[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78

[PIRKW] phosphotransferase 3e-93

[PIRKW] nucleus 2e-74

[PIRKW] calcium 2e-40

[PIRKW] transferase 3e-33

[PIRKW] duplication 2e-32

[PIRKW] tandem repeat 7e-45

[PIRKW] phorbol ester binding 4e-33

[PIRKW] zinc 4e-33

[PIRKW] ion transport 1e-32

[PIRKW] cell cycle control 1e-45

[PIRKW] serine/threonine-specific protein kinase 2e-97

[PIRKW] oncogene 1e-34

[PIRKW] phospholipid binding 2e-32

[PIRKW] autophosphorylation 2e-74

[PIRKW] brain 6e-36

[PIRKW] heterotetramer 8e-38

[PIRKW] mitosis 1e-45

[PIRKW] polymer 5e-41

[PIRKW] magnesium 6e-80

[PIRKW] ATP 2e-97

[PIRKW] polyprotein 1e-34

[PIRKW] alternative initiators 2e-31

[PIRKW] phosphoprotein 2e-74

[PIRKW] apoptosis 8e-38

[PIRKW] cGMP binding 4e-33

[PIRKW] glycoprotein 3e-36

[PIRKW] skeletal muscle 8e-38

[PIRKW] protein kinase 2e-50

[PIRKW] testis 5e-41

[PIRKW] cAMP binding 8e-38

[PIRKW] transforming protein 4e-33

[PIRKW] purine nucleotide binding 7e-52

[PIRKW] calcium binding 7e-45

[PIRKW] alternative splicing 5e-42

[PIRKW] P-loop 7e-52

[PIRKW] lipoprotein 8e-38

[PIRKW] proto-oncogene 4e-33

[PIRKW] segmentation 1e-34

[PIRKW] core protein 1e-34

[PIRKW] muscle 8e-38
 [PIRKW] myristylation 8e-38
 [PIRKW] EF hand 7e-45
 [PIRKW] cell division 3e-49
 [PIRKW] homodimer 1e-32
 [PIRKW] calmodulin binding 5e-42
 [SUPFAM] ribosomal protein S6 kinase II 1e-34
 [SUPFAM] calcium-dependent protein kinase 7e-45
 [SUPFAM] AMP-activated protein kinase 6e-80
 [SUPFAM] protein kinase akt 3e-36
 [SUPFAM] protein kinase SPK1 7e-41
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 5e-42
 [SUPFAM] calmodulin repeat homology 7e-45
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
 [SUPFAM] protein kinase DUN1 6e-36
 [SUPFAM] protein kinase C zeta 4e-33
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
 [SUPFAM] death-associated protein kinase 8e-38
 [SUPFAM] pleckstrin repeat homology 3e-36
 [SUPFAM] ankyrin repeat homology 8e-38
 [SUPFAM] protein kinase homology 8e-99
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 6e-38
 [SUPFAM] protein kinase C zinc-binding repeat homology 4e-33
 [SUPFAM] protein kinase C delta 2e-32
 [SUPFAM] cGMP-dependent protein kinase 3e-33
 [SUPFAM] protein kinase cdrl 1e-45
 [SUPFAM] kinase-related transforming protein 2e-50
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-42
 [SUPFAM] kinase interaction domain homology 7e-41
 [SUPFAM] gag-akt polyprotein 1e-34
 [PROSITE] PROTEIN_KINASE_ATP 1
 [PROSITE] MYRISTYL 3
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 2
 [PROSITE] PROTEIN_KINASE_ST 1
 [PFAM] Eukaryotic protein kinase domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 12.31 %

SEQ MVMADGPRHLQGRPVVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVAVN
 SEG
 lctpEEEECTTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ LEKIYREVQIMKMLDHPHIIKLYQVMTKSMLYLVTEYAKNGEIFDYLANHGRNLNESEAR
 SEG
 lctpE HHHHHHHHHHHHCCTTTBCCCEEEEEETTEEEEEECTTTTBHHHHHHHHCCCCHHHHH

SEQ RKFWQILSAVDYCHGRKIVHRDLKAENLLDNNMNIKIADFQFGNFFKSGELLATWCGSP
 SEG
 lctpE HHHHHHHHHHHHHHCCECCCGGGEETTTTCEEECTTTTTEETT-TTBC-CCCCCG

SEQ PYAAPEVFEGQQYEGPQLDIWSMGVVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPYFM
 SEG
 lctpE GGCCHHHHHCCBCB-HHHHHHHHHHHHHHCCTTTTTHHHHHHHHHCCCCCTTTT

SEQ SEDCEHLIRMLVLDPSKRLTIAQIKHKWMLIEVPVQRPVLYPQEQENEPSIGEFNEQV
 SEG
 lctpE CHHHHHHHHHHTTTTGGGTTTHHHHHHC GG.....

SEQ LRLMHSGLIDQQKTIESLQNKSYNHFAAIYFLLVERLKSRRSSFPVEQRLDGRQRRPSTI
 SEG
 lctpE

SEQ AEQTVAKAQTVGLPVTMHSNMRLLRSALLPQASNVEAFSFPASGCQAEAFMEEECVDT
 SEG
 lctpE

SEQ PKVNGCLLDPPVPLVRKGCQSLPSNMMETSIDEGLETEGEAEEDPAHAFAEFQSTRSGQ
 SEGxxxxxxx.....
 lctpE

SEQ RRRTLSEVTNQLVMPGAGKIFSMNDSPSLDSVDSEYDMGSVQRLNLFEDNPSLKDIML
 SEG
 lctpE


```

SEQ  ANQPSPRMTSPFISLRPTNPAMQALSSQKREVVHNRSPVSFREGRRASDTSLTQGIVAFRQ
SEG  .....
lctpE .....

SEQ  HLQNLARTKGILELNKVQLLYEQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVST
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ  LPASVHPQLSPRQSLETQYLQHLRLQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQK
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ  RLFLQKQSQLQAYFNQMQUIAESSYPQPSQQLPLPRQETPPPSQQAAPPFSLTQPLSPVLEP
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ  SSEQMQYSPFLSQYQEMQLQPLPSTSGPRAAPPLPTQLQQQPPPPPPPPPPRQPGAAPA
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ  PLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQQSDLTGPDGPRSPGLQEAPSSYDPLAL
SEG  xxx.....
lctpE .....

SEQ  SELPGLFDCMLDAVDPQHNGYVLVN
SEG  .....
lctpE .....

```

Prosite for DKFZphtes3_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_15k11.1

HMM_NAME Eukaryotic protein kinase domain

HMM		*YeigRiIGeGsFGtVYkCiWr.TGeIVAIIkkrms.....FlREI	
Query	20	Y I++++G+G+F++V++++R T +VAIKII+K++++ + RE+ YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAPVNLEKIYREV	68
HMM		qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw	
Query	69	QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E QIMKMLDHPHIIKLYQVME-TKSMLYLVTEYAKNGEIFDYLANHGRLNES	117
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKIcDFGLARqM	
Query	118	E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+ ++ EARRKEFWQILSAVDYCHGRKIVHRDLKAENLLDNNMNIKIADFGFGNFF	167
HMM		nnYerMttfCGTPWYMMAPEVIimg.nyYttkVDMWSFGCILWEMMTGep	
Query	168	+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G + KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLDIWSMGVVLYVLVCGAL	215
HMM		PFyddnMemImrIiqrfrfpfWpnCSeElyDFMrwCWnyDPekRPTFrQI	
Query	216	PF++ ++ + + +++ R++++ +SE++ +++R+++ +DP+KR+T+ QI PFDGPTLPILRQRVLEGRFRIPYFMSDCEHLIRMLVLDPKRLTIAQI	265
HMM		LnHPWF*	
Query	266	+H W+ KEHKWM	271

DKFZphtes3_17f10

group: testes derived

DKFZphtes3_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1 CTTCAAGTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51 TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTC AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCC TGAACATAA ATCGTCGGGA
201 AAGATTTTGT CTAGTGAACA CCCTGAATTT CAACCAGCAA CAAACAGCAA
251 TGAAGAAATT GGGCAGAAAA ATATCAGCAG AACTTCATT ACTCAGGAGA
301 CTAAAAAAGG TCCCCCAGTA CTTTGTAGA AGGAGCTTAG GGAAGAAGTA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAG TGGCTTCTGC
401 TGAAATAGAG CCTCCATCAA CAGAAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGCAGGAG ATTCTGTTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTTCTGCTG AAATTTCAGC TCCATCAGCT
651 GAAGAGTCTC CTTCTGTAGA GCTTCTGGCT GAAATTCTGC CTCCATCAGC
701 TGAAGAGTCC CCTTCAGAAG AGCCTCCTGC TGAAATTCCT CCTCCACCAG
751 CTGAAAAGTC TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTGA AGTACAGCCT TTACCAGCTG AGGGCCGCTC
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCTTGT
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCCTAGAGA AGAGGCTCGA
951 GAACCTCAGC TTTCAACAGC TATGGAGACC CTTGCAGAA AGGCTCCTAC
1001 TGAATTTTCA TCTCCATTAC CTAAAGAGAC CACTGCAGAA GAGGCCCTCTG
1051 CTGAAATTTA GCTTCTAGCA GCTACGGAGC CTCCTGCAGA TGAAACTCCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCCCT
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTGC AGATGAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCCTCTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTCAGCT TCTAGCAGCT ACAGAGGCTT CTGCAGAAGA
1451 GGCTCCTGCT GAAGTTCAGC CTCCACCAGC TGAGGAGGCC CCGCTGAAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCGCTGA AGTTCAGCCT CCACCAGCTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGGCCCCGCT GAAGTTCAGC
1651 CTCCACCAGC TGAGGAGGCC CCCTCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCTGT CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCTATAGA
1751 AGAGACCCTT GCTGCAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCTC CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTCAC CTCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGGT AGAGAATGTG TCTACAGAAT
1951 TTCAGTCACC GCAGGTGGCA GGAATTCAG CAGTAAATTT AGGATCGGTT
2001 GTTTTGAAG GTGAAGCAAA ATTTGAAGAG GTTTCAAAAA TCAATTCGTG
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTGAAA
2101 TAGAAAGTGT TTTTCATATA GAATTTAAAC AACGTCCTCC TGAAGTGTAG
2151 TCAGGTGTGA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGGTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGGG ACCCAAGAC TTGGAATACA GGTGGAATAA TGAACAATAA
2301 AAACGTGAGC AGCATAAAAT TACTTGTGTT AATTTCATTC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAATTGCT
2401 TAAATATATC TTCTACAGT AAACCTGTTG ACACGAGTAA AGTTTAATCT
2451 GCAGCATCTT TTTCTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710
Category: similarity to known protein
Classification: unclassified

```
1 MDRSQQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP
101 VLLEDELREE VTPVVQEGS AVKKVASAEI EPPSTEKFPK KIQPLVEEA
151 TAKAEPRPAE ETHVQVPST EETPDAAEAT AVAENSVKQV PPPAEAPLV
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEPPAEI LPPPAEKSPS
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL
301 LPEEAPREEA RELQLSTAME TPAAEAPTEF QSPLPKETTA EEASAEIQLL
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAEETT AEEASAEIQL
401 LAAIEAPADE TPAAEQSPLS EETSAAEAPA EVQSPSAKGV SIEEAPLELQ
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA
501 EEAPAEVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE
551 APSEVQPPPA EEAPAEVQSL PAETPIEET LAAVHSPPAD DVPAAEASVD
601 KHSPPADLLL TEEFFIGEAS AEVSPPPSEQ TPEDEALVEN VSTEFQSPQV
651 AGIPAVKLGs VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTLEIESVFH
701 IELKQRPEEL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P = 7.4e-43

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat
Length = 1,072

HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43
Identities = 185/622 (29%), Positives = 320/622 (51%)

```
Query: 33 SESEIVVISRPDSSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92
      SE +I V+ + + + +E + + + + + E E Q E G + + TS
Sbjct: 436 SEEKIKVVVEKSEKETVIVEEQTEEIQTVEEVTEEDKEAQGEEEEAEEGGEEAATTSP 495

Query: 93 QETKKGPPVLLEDELREEVTPVVQEGSAVKKVASAEIEPPSTEKFPKIQPPLVEEATA 152
      E P + ++EE P + A K + AE + P+ K PA+++ P ++ A
Sbjct: 496 AEEAASPEKETKSPVKEEAKSPAEEKSPAEEK-SPAEEKSPAEVKSPAEVKSPAEEKSPA 554

Query: 153 KAEPRPAEETHVQVPSTEETPDAAEATAVAENSVKQVPPPAEEAP-LVEFFAEIQPPSA 211
      +A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+
Sbjct: 555 EAKS-PAE---VK-SPATVKSPAEEKSPAEEKSPAEEKSPATVKSPGEAKSPAEEKSPA 609

Query: 212 EESP-SVELLAELPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268
      +SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+
Sbjct: 610 VKSPVEAKSPAEEKSPASVKSPGEAKSPAEEKSPAEEKSPATVKSPVEAKSPAEEKSPVT 669

Query: 269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLPPEEAPREEARELQLSTAMETPAE-EAP 327
      V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P
Sbjct: 670 VKS-PAEA---KSPVEVKSPASVKSPSEAKSPAGAKSPAEE-AKS---PVVAKSPAEEKSP 721

Query: 328 TEFQSPPLPKETTAEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHAQVQS--- 383
      E + P ++ AE S A + PA+ ++PAEA+SP+ E S E+A + V+
Sbjct: 722 AEAKPPAEAKSPAEEKSP-----AEAKSPAEEKSPAEEKSPV-EVKSPKAKSPVKEGAK 775

Query: 384 PLAEETTAEASAEIQLLAIEAPAD-ETPAEAQSPLSEET-SAEAPAE-EVQSPSAKGV 440
```


LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK
 Sbjct: 776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTP 833
 Query: 441 SIEEA--PLELQPPSGEETTA--EASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494
 + EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +
 Sbjct: 834 AKEEAKRPADIRSPEQVKSPEKEAKSPEKEETRTKVPAPKKEEVKSPVEEVKAKEPPKK 893
 Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553
 V+ P EV+ +EAP E Q P AEE + P +++P E + EEA
 Sbjct: 894 VEEKTPATPKTEVKESKKDEAPKEAQKPKAEKEPLTEKP--KOSPGEAKK---EEAKE 948
 Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAAEASVD-KHS 603
 + P EE PA++ ++ P AE+ +E + P ++VPA D K
 Sbjct: 949 KKAAPAEETPAKLGVKKEAKPKEADAKAKEPSKPKSEKEKPKKEVPAAPEKDDTKKEE 1008
 Query: 604 PPADLLLTEEFFIGEASAESVSP--PSEQT-PEDEALVENVSTEFQSPQ 649
 + EE P +A A+ P E + P+ E ++ ST+ + Q
 Sbjct: 1009 KTTESKKPEEKPKMQAKAKEEDKGLPQEPSKPKTEKAESSTSDQKDSQ 1057

Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42
 Identities = 184/628 (29%), Positives = 310/628 (49%)

Query: 18 IPPVEKVDKEQQTYFSESEIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74
 I VEK +KE ++E + ++ + E+ + + G+ A+ P + A
 Sbjct: 440 IKVVEKSEKETVIVEEQTEETIQTVEETVEEDKEAQGEEEEAEEGGEEAATTSPPAEEA 499
 Query: 75 TNSNEEIGQKNISRTSFTQETKKGPPVLLEDELREEVTPVQVQEGSAVKKVASAEIEPPS 134
 + +E + + + + K P E + E P + A K + AE + P+
 Sbjct: 500 ASPEKET-KSPVKEEAKSPAEAKSPA---EAKSPAEAKSPA EVKSPA EVK-SPA EAKSPA 554
 Query: 135 TEKFPKAIQPLVEEATAKAEPRAEETHVQVQ-PSTEETPDAAEATAVAENSVKVPQPP 193
 K PA+++ P ++ A+A+ ++ +V+ P+T ++P + A A++ +V+ P
 Sbjct: 555 EAKSPA EVKSPATVKS PAEAKSPA EAKSPA EVKSPATVKS PGEAKSPA EAKSPA EVKSPV 614
 Query: 194 AEEAPL-VEFPAEIQPPSAEESPS-VELLAELPPSAEESPS-EPPAEILPPPAEKSPS 250
 ++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+
 Sbjct: 615 EAKSPA EAKSPA SVKSPGEAKSPA EAKSPA EVKSPATVKS PVEAKSPA EVKSPVTVKSPA 674
 Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPEEAPR 307
 + E++SP++ K+P E + P A+ E ++P + P ++ AE +P ++P
 Sbjct: 675 EAKSPA EVKSPA SVKSPSEAKSPA GAKSPA EAKSPA VVAKSPA EAKSPA EAKPPAEAKSPA 734
 Query: 308 EEARELQLSTAME--TPAE-EAPTEFQSP----LP-KE---TTAEASAEIQLLAATE-- 354
 E + + E +PAE ++P E +SP P KE + AE S E E
 Sbjct: 735 EAKSPA EAKSPA EAKSPA EAKSPA EVKSPA EAKSPA VKEGAKSLAEAKSPA EAKSPA VKEEI 794
 Query: 355 -PPAD-ETPAEARSPLSEET-SAEAAH-EVQSP LAEETTAEEAS--AEIQLLAIEAPA 408
 PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ +++PA
 Sbjct: 795 KPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPEQVKS 854
 Query: 409 DETPAEAQSPLESEETSAEE-APA--EVQSPSAKGVSEIEAPLELQPPSGEETTAEEASAA 465
 E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A
 Sbjct: 855 KE---EAKSPEKEETRTKVPAPKKEEVKSP-----VEEVKAK-EPPKKVE---EETKPA 901
 Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525
 E+ +EAP E Q P AEE + P +++P E + A+E A P E
 Sbjct: 902 TPKEVKESKKDEAPKEAQKPKAEKEPLTEKP--KOSPGEAKKEEAKKAAA---PEE 956
 Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581
 E PA++ + P E+A P++ PSE + P EE PA + +E E+
 Sbjct: 957 ETPAKLGVKKEAKPKEADAKAKEPSK--PSEKEKPKKEVPAAPEKDDTKKEETTESK 1014
 Query: 582 AAVHSPADDVPAAEASVDKHSPPADLL-LTEEFFIGEASAESVSPSPSEQTPEDEA 636
 P EE DK P TE+ ++ + PSE+ PED+A
 Sbjct: 1015 KPKEPKMQAKAKEE---DKGLPQEPSKPKTEKAESSTSDQKDSQPSEKAPEDKA 1067

Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36
 Identities = 162/540 (30%), Positives = 275/540 (50%)

Query: 135 TEKFPKAIQPLVEEATAKAEP-----PAEETHVQVQVQSTEETPDAAEATAVAENSVKV 189
 TE P KI P + K+E + +E+ V V+ TEE E T E +
 Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEETIQTVEETVE--EEDKEA 474
 Query: 190 QPPPAEEAPLVEFPAEIQPPSAEESPSVELLAELPPSAEE--SPSE-EPPAEILPPPAE 246
 Q EEA A P AEE+ S E E P EE SP+E + PAE P
 Sbjct: 475 QGEEEEAEEGGEEAATTSPPAEAAASPE--KETKSPVKEEAKSPA EAKSPA EAKSPA EAKSPA 532
 Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
 KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + ++P
 Sbjct: 533 KSPA-----EVKSPA EVKSPA EAKSPA-PAEA---KSPA EVKSPA TVKSPA EAKSPA EAKSPA 583

Query: 307 REEARELQQLSTAME--TPAE-EAPTEFQSPLPKETTAEAS-AEIQLLAATEPPAD-ETP 361
 E + + E +PAE +SP+ ++ AE S A ++ + PA+ ++P
 Sbjct: 584 AEVKSPATVKSPGEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 643

Query: 362 AEARSPLSEETSAE-EAHAEVQSPLAEETTAEASAEIQLLAAIEAPAD-ETPAEQSPL 419
 AE +SP + ++ E ++ AEV+SP+ ++ AE A + ++ +++PA ++P+EA+SP
 Sbjct: 644 AEVKSPATVKSPVEAKSPAEEKSPVTVKSPA-EAKSPVE---VKSPASVKSPSEAKSP- 697

Query: 420 SEETSAEEAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAE 478
 + ++PAE +SP AK + ++P E +PP+ ++ AE S A A + A A+
 Sbjct: 698 ----AGAKSPAEEKSPVAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 749

Query: 479 APAEVQPPPAEEAPAEVQPPPAEEAP--AEVQPPPAEEAP--EVQPPPAEEAPAEVQPP 534
 +PAE + P ++P + + P E A AE + P ++P E++PP ++P + + P
 Sbjct: 750 SPAEAKSPVEVKSPKAKSPVKEGAKSLAEAKSPKAKSPVKEEIKPPAEVKSPKAKSP 809

Query: 535 PAEEAPAEVQPPPAEEAPSEVQPPPAEEA--PAEVQSLPAEETPIETLAAVHSPADDV 592
 EEA + + + E + P EEA PA+++S ++P +E SP ++
 Sbjct: 810 MKEEAKSPEKAKTLOVKSPEAKTPAKEEAKRPADIRSPQVKSPEAKE---AKSPEKEET 866

Query: 593 PAEEASVDKHS--PPADLLLTEEFFIGEASAEVSPFPSEQTPDEALVENVSTEFQSPQV 650
 E+ + K P + + +E P + E P + +T E+ + E Q P+
 Sbjct: 867 RTEKVAPKKEEVKSPVEVKAKEPP--KKVEEEKTPATPKTEVKESKKDEAPKEAQKPKA 924

Query: 651 AGIPAVKLGSVVLEGEAKFEVSK 674
 + GEAK EE +
 Sbjct: 925 EEKEPLTEKPKDSPGEAKKEEAK 948

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34
 Identities = 123/390 (31%), Positives = 213/390 (54%)

Query: 308 EEARELQQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPA---EA 364
 E+ E+Q++ E EE E Q +E AEE E A T PPA+E + E
 Sbjct: 455 EQTEEIQVT---EEVTEEDKEAQGE---EEEEAEEGGEEA---ATTSPPAEEAASPEKET 506

Query: 365 RSPLSEETSAEEAHAEVQSPLAEETTAEAS-AEIQLLAAIEAPAD-ETPAEQSPLSEE 422
 +SP+ EE + AE +SP ++ AE S AE++ A +++PA+ ++PAEA+SP +
 Sbjct: 507 KSPVKEEAKSP---AEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 563

Query: 423 TSAE-EAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAEAP 480
 + A ++PAE +SP+ AK + ++P ++ P GE + EA + ++ + EA ++P
 Sbjct: 564 SPATVKSPAEEKSPAEEKSPAEEKSPAATVKSP-GEAKSPAEEKSPAEEKSPAEEKSPA 619

Query: 481 AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP 540
 AE + P + ++P E + P ++PAEV+ P ++P E + P ++P V+ P ++P
 Sbjct: 620 AEAKSPASVKSPGEAKSPAEEKSPAEEKSPAATVKSPVEAKSPAEEKSPAEEKSPA 679

Query: 541 AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIETLAAVHSPAD-DVPAEEASV 599
 EV+ P + ++PSE + P ++PAE +S ++P E A PPA+ PAE S
 Sbjct: 680 VEVKSPASVKSPSEAKSPAGAKSPAEEKSPVAKSPAEEKSPAEEKSPAEEKSPA 739

Query: 600 DKHSPADLLLTEEFFIGEASAEVSPFPSEQTPDEALVENVSTEFQSPQVAGIPAVKLG 659
 + PA+ E ++ EV P ++P E +++ E +SP+ A P VK
 Sbjct: 740 AEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 792

Query: 660 SVVLEGEAKFEVSKINSVLKDLNTNDGQAPTLEIES 697
 + E K E +K S +K+ + + +A TL+++S
 Sbjct: 793 EIKPPAEVKSPKAK--SPMKEEAKSPE-KAKTLDVKS 827

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18
 Identities = 124/420 (29%), Positives = 199/420 (47%)

Query: 252 ELLGEIRSPAQAQAPIEVQPLPA-----EGALEEAPAKVEPPTVEETLAEVQPLLPPEAP 306
 ELLG+I+ A +A + + A AL E A++E TV+ TL +
 Sbjct: 236 ELLGQIQGCGAAQAQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEWFRVRLDR 295

Query: 307 REEARELQQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPAEARS 366
 EA ++ + AM + EE TE++ L TT E++ L +T+ + +E
 Sbjct: 296 LSEAAKVN-TDAMRSAQEEI-TEYRRQLQARTT-----ELEALKSTKESLERQRSELED 347

Query: 367 PLSEE-TSAEEAHAEVQSPLAEETTAEASA--EIQLLAAIEAPAD-ETPAEQSPLSEE 422
 + S ++A ++ + L T E A+ E Q L ++ D E A + EE
 Sbjct: 348 RHQVDMASYQDAIQQLDNEL-RNTKWEMAAQLREYQDLLNVKMLDIEIAAYRKLLGEE 406

Query: 423 TSAEEAPAEV-----QSPS-AGVSIIEAPLELQPPSGEETT-AEEASAAIQLLA-A 471
 P+ + PS + + ++ E +++ S +ET EE + IQ+
 Sbjct: 407 CRIGFGSPFSLTEGLPKIPSMSTHIKVKSEEKIKVVEKSEKETVIVEEQTEEIQTVEEV 466

Query: 472 TEASAEAPAEVQPPPAEEAPAEVQPP--PAEEAPAEVQPPPAEEA--PAEVQPPPA 524
 TE +EA E + AEE E PPAEEA + E + P EEA PAE + P
 Sbjct: 467 TEEEDKEAQGE-EEEEAEEGGEEAATTSPPAEEAASPEKETKSPVKEEAKSPAEEKSPA 525

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18
Identities = 115/364 (31%), Positives = 166/364 (45%)

Report for DKFZphtes3_17f10.3

```

SEQ      MDRSQQTSRTGYWTMMNIPPVEKVDEKQOTYFSESEIIVVISRPDSSSTKSKEDALKHKSS
SEG      .....
PRD      cccccccccccccccccccceehhhhhhhccccceeeecccccccccchhhhhhhccc

SEQ      GKIFASEHPEFPQATNSNEEIGQKNISRTSFQTETKGGPPVLEDELREEVTVPVVQEGS
SEG      .....
PRD      cceeeccccccccccccccccccccccccccccceeeecccccchhhhhhhhheeeecccc

SEQ      AVKKVASAEIEPFPSTEKFPKIQPPLVEEATAKAEPRAEETHVQVPQSTTEETPDAAEAT
SEG      .....xxxxxxxxxxxx
PRD      chhhhhhhccccccccccccccccchhhhhhhccccccceeeecccccccccchhhh

SEQ      AVAENSVKVQPPPAEAPLVEFPAEIQPPSAEESPSVELLAEILPPSAEESPSPEEPPAEI
SEG      xxxxx.....xxxxxxxxxxxxxxxxxxxx
PRD      hhhhccccccccccccceeeccccccccccccccchhhhhcccccccccccccccc

SEQ      LPPPAEKSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEFPPTVEETLAEVQPL
SEG      xxxxxx.....xxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccccccccccccccccchhhhhccccccccccccchhhhhhhhhc

SEQ      LPEEAPREEARELQLSTAMETPAEEAPTEFQSPPLKETTAEASAEIQLLAATEFPADET
SEG      xxxxxxxxxxxxxxxxx.....xxxxxxxxxxxx
PRD      cccccchhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccc

```


(No Pfam data available for DKFZphtes3_17f10.3)

DKFZphtes3_17117

group: metabolism

DKFZphtes3_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```

1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51 GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGCACGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTGCTGT CTGTCTCTTT CTTCCACACG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACACAG ACCGGTTCAT CCTCTCCAGG GGACATGCTG
251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT.
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGGAGA GACACCTTAC
351 CCCGCGATTG CCGTTTGTG ACGTGGCAAC AGGGTCCCTA GGTCAGGGAT
401 TAGCTACTGC ATGTGGAATG GCTTATACTG GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCCTT TATGGGAGAT GCGGAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGCTT TTGCCTCCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
601 GAGCATGGCG CAGACATCTA CCAGAATTGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGGTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATA CAAAATCGCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCTTA
951 CAAAGTTGGT GACAAGATAG CTACTCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAGCTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAAGTCCA CCTTTTCTGA GATATTCAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAGTGTG GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTTCGATCA GCTCCGAATG GGAGCCATTT CTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGCCTGGGAG GATCTAGCCA TGTTCCGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTTCGAAC AGCCAACCAG
1451 AAAGTGCAGT TATTTATACC CCACAAGAAA ATTTTGAGAT TGGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTCACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTCTCTG CCGTGTCTATC GACCCATTTA CCATTAAACC CCTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GCGGCGCGAG TTATCAGAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTTGTTTCATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAAAACCTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1901 TTATTCTTAA AAAGTCAAGT CTATTGGCTT TGGCCCCAAA GCACTGGTAT
1951 CTTTGTATTA AATTCTATGT TATTGTCACA AAACCATTTT TTATACCTAT
2001 ACAGTTGTAC TGTTCCTTTT AAAGCAAAGC CATTTAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2101 TTAAGTCTGAG TTATTAATGT GGATTTTAAA ATTGTAAGCA ATAGAATAGG
2151 AAATAAACA ACTACCTAAT ACAAATATTT CTGATAAGAC TACAAATATC
2201 TGACTGAGCT GGGGATTAAG GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCTCTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTGT AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTACAG ATAATGTTT ACTGCAGTTG CCTTGGAAAT TCCTCCAAAG
2401 TTTGCCCTCA TCTCTCTCT ACAGTTTGGG GGTGATGGTG CAGCAGTGGA
2451 ACATCTCTTG ATGCACCACA CTACTTGTGT TCTGTGAAGT GATGAAAGTA

```


2501 TAACTGGTTC TAGTTTGCAC ACTACACACA TAGTTTGTG AAGCTTCAGA
 2551 AATGTTTTTT CTTTCTCTG TGGCCAAACC AGTTTGTAA TCTGATTATA
 2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAAGCAT TAAAAATCA
 2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

BLAST Results

No BLAST result

Medline entries

96214928:
 Amplification of the transketolase gene in desensitization-resistant mutant
 Y1 mouse adrenocortical tumor cells.

99123875:
 Properties and functions of the thiamin diphosphate dependent enzyme transketolase.

Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626
 Category: strong similarity to known protein
 Classification: Metabolism
 Prosite motifs: ATP_GTP_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS
 51 VLFFHTMKYK QTDPEHPDND RFILSRGHAA PILYAAWVEV GDISESDLLN
 101 LRKLHSDLER HPTPLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV
 151 FCLMGDGESS EGSVWEAFAP ASHYNLDNLV AVFDVNLGQ SGPALEHGA
 201 DIYQNCCEAF GWNTYLVDPGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR
 251 GIPNIEDAEN WHGKVPVKER ADAIVKLIES QIQTENLIP KSPVEDSPQI
 301 SITDIKMTSP PAYKVGDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST
 351 FSEIFRKEHP ERFIECIIAE QNMVSVALGC ATRGRTIAFA GAFAAFFTRA
 401 FDQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRISPNCTVF
 451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR
 501 HGVNDKVTVI GAGVTLHEAL EAADHLSQQG ISVRVIDPFT IKPLDAATII
 551 SSAKATGGRV ITVEDHYREG GIGEAUCAAV SREPDIIVHQ LAVSGVPQRG
 601 KTSLELDMEF ISTRHIIAAV TLTLMK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17117, frame 1

SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).. N = 1,
 Score = 2222, P = 2.5e-230

SWISSPROT:TKT_RAT TRANSKETOLASE (EC 2.2.1.1) (TK).. N = 1, Score =
 2202, P = 3.3e-228

TREMBL:RN09256_1 product: "transketolase"; Rattus norvegicus
 Sprague-Dawley transketolase mRNA, complete cds.. N = 1, Score = 2202,
 P = 3.3e-228

SWISSPROT:TKT_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK).. N = 1, Score =
 2200, P = 5.3e-228

>SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).
 Length = 623

HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230
 Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFFHTM+YK DP +
6 KPQQKQLQALKDTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYAANVEVDISEDLLNLRLKLSDLERHPTPLPFVDVATGSL 126
P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLDGHPVPKQAFDVTATGSL 125

Query: 127 GQGLGTACGMAYTGKYLKASYRVFCLMGDGESEGSVWEAFASHYNLONLVAVFDVN 186
GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA Y LDNLVA+FD+N

Sbjct: 126 GQGLGAACGMAYTGKYFDKASYRVYCLMGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGSPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246
RLGQS PAPL+H DIYQ CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQSDPAPLQHQVDIYQKRCEAFGWHTIIVDGHVSVEELCKAFGQA---KHQPTAIIAKT 242

Query: 247 FKGRGIPNIEDAENWHGKPVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQISITDIK 306
FKGRGI IED E WHGKP+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEISYQVQSKKKILATPPQEDAPSVDIANIR 302

Query: 307 MTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHPERFIEC 366
M +PP+YKVGDKIAT+K YGLALAKLG A++R+I L GDT NSTFSE+F+KEHP+RFIEC

Sbjct: 303 MPTPPSYKVGDKIATRKYGLALAKLGHASDRIIALOGDTKNSTFSELFKKEHPDRFIEC 362

Query: 367 IIAEQNMVSVALGCATRGRTIAFAGFAFAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426
IAEQNMVS+A+GCATR RT+ F FAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSIIVGATRDRTVPFCSTFAFFTRAFDQIRMAAISSEININLIGSHCGVSIG 422

Query: 427 EDGVSQMALEDLAMFRSIPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPETAIVIY 486
EDG SQMALEDLAMFRS+P TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAMFRSVPMSSTVFYPSDGVATEKAVELAANTKGICFIRTSRPNIAIYS 482

Query: 487 PQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEAADHLSQQGISVRVIDPFTIKPLDA 546
E+F++GQAKVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIKPLD

Sbjct: 483 NNEDPQVGGQAKVVLKSKDDQVTVIGAGVTLHEALAAAEESLKKDKISIRVLDPFTIKPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYEGGIGEAUAVSREPDIIVHQLAVSGVPQGRKTSELL 606
I+ SA+AT GR++TVEDHY EGGIGEAU AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYEGGIGEAUAAVVEGPGVTVTRLAVSQVPRSGKPAELL 602

Query: 607 DMFGISTRHIIAAV 620
MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKFZphtes3_17117, frame 1

Report for DKFZphtes3_17117.1

[LENGTH] 626
[MW] 67877.52
[pI] 5.90
[HOMOL] SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0
[FUNCAT] m outer membrane and cell wall [M. jannaschii, MJ0681] 3e-48
[FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17
[FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09
[FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
[BLOCKS] BL00801F
[BLOCKS] BL00801E
[BLOCKS] BL00801D Transketolase proteins
[BLOCKS] BL00801C Transketolase proteins
[BLOCKS] BL00801B Transketolase proteins
[BLOCKS] BL00801A Transketolase proteins
[SCOP] dltka2 3.28.1.2.1 Transketolase, C-terminal domain 1e-21
[EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11
[EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10
[EC] 2.2.1.1 Transketolase 0.0
[EC] 2.2.1.3 Formaldehyde transketolase 1e-20
[PIRKW] transferase 0.0
[PIRKW] flavoprotein 2e-07
[PIRKW] Calvin cycle 1e-40
[PIRKW] heterotetramer 2e-07


```

[PIRKW]      pentose phosphate pathway 0.0
[PIRKW]      magnesium 1e-40
[PIRKW]      thiamine pyrophosphate 0.0
[PIRKW]      oxidoreductase 7e-12
[PIRKW]      fatty acid biosynthesis 4e-10
[PIRKW]      mitochondrion 2e-07
[PIRKW]      peroxisome 1e-20
[PIRKW]      homodimer 1e-40
[SUPFAM]     pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
[SUPFAM]     pyruvate dehydrogenase (lipoamide) beta chain 7e-12
[SUPFAM]     ferredoxin 2[4Fe-4S]-related protein 8e-47
[SUPFAM]     thiamine pyrophosphate-binding domain homology 0.0
[SUPFAM]     pyruvate dehydrogenase (lipoamide) 6e-08
[SUPFAM]     ferredoxin 2[4Fe-4S] homology 8e-47
[SUPFAM]     hypothetical protein C2814 2e-21
[SUPFAM]     transketolase 0.0
[PROSITE]    ATP_GTP_A 1
[PFAM]       Transketolase
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY 3.04 %

```

```

SEQ      MMANDAKPDVKTQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYK
SEG      .....
lngsB    .....HHHHHHHHHHHHCCCCHHHHHHHHHHHHHH-HHCCCT

SEQ      QTDPEHPDNDRFILSRGHAAPILYAAMVEVGDISDNLNLRKLHSDLERHPTPLRPFVD
SEG      .....
lngsB    TTTTTTTTCEEETGGGHHHHHHHHHHCTTCHHHHHTTTTTTTTTTTTTTTTTTTTC

SEQ      VATGSLGQGLGTACGMAYTGKYLKASYRVFCLMGDGESSEGSVWEAFASFASHYNLDNLV
SEG      .....
lngsB    CCCCTTTHHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHHHHHHHHHHHHHHHCTTTEE

SEQ      AVFDVNRLQGSGPALEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPT
SEG      .....
lngsB    EEEEECCETEETEGGGCCCCCHHHHH-HHHCCEEETTTTHHHHHHHHHHHHHHTTTTCE

SEQ      AIVAKTFKGRGIPNIEDAENWHGKVPVKERADAIVKLIESQIQTNENLIPKSPVEDSPQI
SEG      .....
lngsB    EEEECTTTTTTCCHHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ      SITDIKMTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
SEG      .....
lngsB    HHHHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHHHTTTTTEEEETTTTHHHHCCTTCECCG

SEQ      ERFIECIIAEQNMVSVALGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSH
SEG      .....
lngsB    GCEEETTTTHHHHHHHHHHHHHHTTTTEEEEGGGGGGGHHHHHHHHHHCTTTEEEEC

SEQ      CGVSTGEDGVSQMALEDLAFERSIPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPE
SEG      .....
lngsB    CCGGGTTTTTTTTTCCHHHHHHHHCTTTTEEECCCHHHHHHHHHHTTTTCEEECCCCCB

SEQ      TAVIYTPQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEAADHLSQQGISVRVIDPFT
SEG      .....
lngsB    CCTTTTCHHHHHCC-CEEETTTTTEEEECCHHHHHHHHHHHHHHHHHHCCCEEE....

SEQ      IKPLDAATIISAKATGGRVITVEDHYREGGIGEAACAASVREPDIIVHQLAVSGVPQRG
SEG      .....
lngsB    .....

SEQ      KTSSELLDMFGISTRHIIAAVTLTLMK
SEG      .....
lngsB    .....

```

Prosites for DKF2phtes3_17117.1

PS00017 595->603 ATP_GTP_A PDOC00017

Pfam for DKF2phtes3_17117.1

HMM_NAME Transketolase

HMM *vNtIRiLaMDAVEKANSQHPGaPMGMAPMAHVVLWqrMMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL++++M++++DP P+ ANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYKQTDPEHPD	68
HMM		RDRFVLSNGHaCMLLYsMWHLyGYDmpMWDLkQFRQWHSrTPGHPEIghT +DRF+LS GHA+++LY+ W + G +++++DL+++R++HS++ +HP ++	
Query	69	NDRFILSRGHAAPILYAANWVEVGD-ISESDLLNLRKLHSDLERHPTPRLP	117
HMM		PGVEVTGPLGQGIaNaVWMAIAERnLAATYNRPGFDI fDHYTYCFMGDG ++ +V+TG+LGQG++ +++++Y++++ D++++++C+MGDG	
Query	118	FV-DVATGSLGQLG-----TACGMAYTGKYLDKASYRVFCLMGDG	157
HMM		CLMEGISWEACSLAGHMqLGNWiaFYDDNrISIDGdTdIWfQEdtYakRF + +EG++WEA ++A+H++L+N++A +D NR++++G++++ + D+Y+ +	
Query	158	ESSEGSVWEAFAFASHYNLDNLVAVFDVNRQLGQSGPAPLEHGADIYQNC	207
HMM		EAYGWHVIEVEnDGHdV eEiCaIEeAKaekDRPTLIiCRTVIGYGS PNk EA+GW++ +V DGHdVE++C A+ +A +K++PT+I ++T++G+G+PN	
Query	208	EAFGWNTYLV--DGHdVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI	255
HMM		QGTdHWHGAPLGeD* ++ + WHG+P +++	
Query	256	EDAENWHGKVPVKE 269	
HMM		*PqWePnddkIATRKASQqaleaiGPALPEfWGGsADLTpSNLTiWKGMv P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S++++++ ++	
Query	311	PAYKV-GDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKE	358
HMM		WFMPPSISTDCynGNWsGRYIHYGIREHgMgAIMNGIALHGqNFRPYGGT + + R+I++ I+E++M++++ G+A++G+ +++++ G	
Query	359	H-----PERFIECIIAEQNMVSVALGCATRGR-TIAFAGA	392
HMM		FMMfyDYARPAIRMAALMelPVIWVTHDSIGLGEDGPTHQPVEHLAHR F++F+++A++++RM A++ +++++H++++ GEDG +++++E+LA+FR	
Query	393	FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSQMALEDLAMFR	442
HMM		aIPNMsvVRPCDgNETayAWylAvERehTPtiLILSRQNLPLQLerNPrqf +IPN +V++P+D+ T+ A YLA++++ +++++S ++ +++++ P +	
Query	443	SIPNCTVFYPSDAISTEHAIIYAANTKGM-CFIRTSQPETAVIYT-PQEN	490
HMM		ekvaRGGYVLkDmdnePDVILIATGSEMELavaAAKlLadEGikaRVVSM +++++++V + + + V++I++G++++A++AA+ L+ +GI +RV+++	
Query	491	FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQQGISVRVIDP	538
HMM		PCTeWFD.....kQDeEYReSVLPDhVPqRVaVEmGvtWCWYKYVGqq +++++D + +++++R +++DH++ ++++++V ++ +++ +	
Query	539	FTIKPLDAATIISAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDIL	587
HMM		GaIfGMNrFGESSGKAPpevLYkMFGFTPENI* + +++ +++ ++ +L+ MFG+ +I	
Query	588	VHQLAVSGVPQR---GKTSELLDMFGISTRHI 616	

DKFZphtes3_17n12

group: transcription factors

DKFZphtes3_17n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucine-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```
1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGTCTCAT GTAACAATAG TAGATTGTTT TTTTCCTAA TATTTCTAGC
101 CAGGCCCTGAA GTCAGGTGAT GGAACAAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GGTGGAGGAA GGAAAGAAGA AAAATGGGAA GAATGTCTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCCTGTGC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTTAACCTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
301 GTGGCCTCCC ATCTGCCTCT GCACCCATA ATGCACAACA AACCTCACTC
351 TGAGGAGGTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
401 ACAGCGTTCT GTCATCTCAG CAAAGAATGG AATCAGAGAA TAATAAGTTA
451 TGTTCCTTAT ATTCTTCCG AAATACCTCT ACCTCACCAC ATAAGCCTGA
501 CGAAGGGGAG CGGGACCGTG AGATAATGAC CAGTGTACT TTTGGAACCC
551 CAGAGCGCGC CAAAGGGAGT CTGCGGATG TGGTGGACAC ACTGAAACAG
601 AAGAAGCTTG AGGAAATGAC TCGGACTGAA CAAGAGGATT CCTCCTGCAT
651 GGAAAAAATA CTTTCAAAAG ATTGGAAGGA AAAAATGGAA AGACTAAATA
701 CCAGTGAAC TCTTGGAGAA ATTAAAGGTA CACCTGAGAG CCTGGCAGAA
751 AAAGAACGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTACGGGA
801 GCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA GCGTCACAAA
851 TTGAGAAACA ACGGCAGCAA ATGGACCTTG CTCGCCAACA GCAAGAACAG
901 ATTGCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCCCTCG CTCATGATCC
1001 CAATTTTCC ACATGACCG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGGATTCC TCTTCCCCC TGAATAACA TACAAACCA GTGATACTA
1101 CCCCAGTACG TTCATTCCAT CAACAATGGC AGTGTCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTAC CTGGAGCAA GATGCCATCA ACTCCACAGC CACCAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCACAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGACAGC AGAGCCTGTA AAGTCCCCAA CGTCTCCAC
1401 CCAGAACCTC TCCCAGCCA GCAAACCAAG CCCTGTCAAT CTGCCAAACA
1451 AAAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAAAGTCA ACACCAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAGCCCAT TCAGGAGCGG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTG
1751 AGAATTTGGG GCCCCAGTTA ACGGGAAAGT CAAATGAAGA TGGAAAACCTG
1801 GGCCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGCAG CTAAACTACA GCAGTATTAT TGTGGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCAGCAG TCTACAGGGA CGCCCGCGGC
1951 CGTGCCAGCA GCGAGCCACA CATTAAGCGA CCAATGAATG CATTCATGGT
2001 TTGGGCAAAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCGGACATGC
2051 ATAATCCCAA CATTAGCAAA ATCTTAGGAT CTCGCTGGAA ATCAATGTCC
2101 AACCAGGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAACCCCGA CCGAAACGCA
2201 CCTGCATTGT TGATGGCAAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGCAG TTCTTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACCA CAGGAACAGG TGTGTGTAT CCTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGCTCT
```



```

2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAAATG TATGATGACT ATGAAGATGA CCCCAAATCA
2551 GACTATAGCA GTGAAATGA AGCCCGGAG GCTGTCACTG CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCA
2651 ACAAGAGATT ATTAAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

BLAST Results

No BLAST result

Medline entries

95311974:
A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:
The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804
Category: strong similarity to known protein

```

1 MGRMSSKQAT SPFACAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESENNKLC SLYSFRNTST
101 SPHKPDEGSR DREIMTSVTF GTPERRKGS LADVVDLTKQ KLEEMTRTEQ
151 EDSSCMEKLL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMITQ
201 LISLREQLLA AHDEQKKLAA SQIEKQRQQM DLARQQEQEI ARQQOQLLQQ
251 QHKINLLQQQ IQVQGHMPPL MIPIFPHDQR TLAAAAAAQQ GFLFPFGITY
301 KPGDNYPVQF IPSTMAAAAA SGLSPLQLQQ LYAAQLASMQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSPVTQVKD EAAAQPLNLS SRPKTAEPVK
401 SPTSPTQNLF PASKTSPVNL PNKSSIPSPI GGS LGRGSSL GKWKSQHQEE
451 TYELDILSSL NSPALFGDQD TVMKAIQEAR KMREIQIREQ QQQQPHGV DG
501 KLSSINNMGL NSCRNEKERT RFENLGPQLT GKSNEGKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQYYC WPTGGATVAE ARVYRDAGR ASSEPHIKRP
601 MNAFMVWAKD ERRKILQAFP DMHNSNISKI LGSRWKSMN QEQOPYEEQ
651 ARLSKIHLEK YPNYKYPKP KRTCIVDGKK LRIGEYKQLM RSRROEMRQF
701 FTVGQQPQIP ITTGTGVVYP GAITMATTPP SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMEMY DDYEDDPKSD YSSENEAPEA
801 VSA

```

BLASTP hits

Entry MMSOXLZ2_1 from database TREMBL:
product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds.
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:
SOX-LZ - rainbow trout
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:
SOX6 protein - mouse
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330_1 from database TREMBL:
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604_1 from database TREMBL:
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281

Alert BLASTP hits for DKFZphtes3_17n12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n12, frame 1

Report for DKFZphtes3_17n12.1

[LENGTH] 804
 [MW] 89332.69
 [PI] 6.97
 [HOMOL] TREMBL:MMSOXL22_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0

[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04
 [SCOP] dlhmf_1.20.1.1.1 HMGI, fragments A and B [rat/hamster (Rattus)] 1e-13
 [SCOP] dllefa_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mous] 4e-15
 [SCOP] dlhrya_1.20.1.1.4 SRY [Human (Homo sapiens)] 7e-17
 [PIRKW] DNA binding 4e-94
 [PIRKW] T-cell receptor 4e-07
 [PIRKW] leucine zipper 1e-38
 [PIRKW] alternative splicing 2e-07
 [PIRKW] transcription factor 4e-16
 [PIRKW] transcription regulation 1e-12
 [SUPFAM] HMG box homology 0.0
 [SUPFAM] unassigned HMG box proteins 4e-94
 [PROSITE] ATP_GTP_A 1
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] HMG (high mobility group) box
 [KW] Irregular
 [KW] 3D
 [KW] LOW COMPLEXITY 13.81 %
 [KW] COILED_COIL 3.48 %

SEQ MGRMSSKQATSPFACAAADGEDAMTQDLTSREKEEGSDQHVASHLPLHPIMHNKPHSEELP
 SEG
 COILS
 lnhm-

SEQ TLVSTIQDADWDVSLSSQQRMESENNKLCSLYSFRNTTSPHKPDEGSRDREIMTSVTF
 SEG
 COILS
 lnhm-

SEQ GTPERRKGSADVVDTLKQKKLEEMTRTEQEDSSCMEKLLSKDWKEKMERLNTSELLGEI
 SEG
 COILS
 lnhm-

SEQ KGTPESLAERQQLSTMITQLISLREQLLAHDEQKLAASQIEKQRQCMDLARQQEQEI
 SEG
 COILS
 lnhm-

SEQ ARQQQQQLLQQQHKNLLQQQIQVGHMPPLMIPIFPHDQRTLAAAAAQQGFLFPPGITY
 SEG
 COILS
 lnhm-

SEQ KPGDNPVQFIPSTMAAAAASGLSPLQLQQLYAAQLASMQVSPGAKMPSTPQPPNTAGTV
 SEG
 COILS
 lnhm-


```

COILS .....
lnhm- .....

SEQ      SPTGIKNEKRGTSPTVQVKDEAAQPLNLSSRPKTAEPVKSPSTSPTQNLFPASKTSPVNL
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PNKSSIPSPIGGSLGRGSSLGKWKSQLHQEETYELDILSSSLNSPALFGDQDTVMKAIQEAR
SEG      .....
COILS    .....
lnhm-    .....

SEQ      KMRREQIQREQQQQPHGVGDKLSSINMGLNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PGVIDLTRPEDAEGSKAMNGSAAKLQYYCWPTGGATVAEARVYRDARGRASSEPHIKRP
SEG      .....
COILS    .....
lnhm-    .....CCC

SEQ      MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEKQPYEEQARLSKIHLEK
SEG      .....x
COILS    .....
lnhm-    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      YPNYKYKPRPKRTICVDGKKLRIGEYKQLMRSRRQEMRQFFTVGQQPQIPITTTGTGVVYP
SEG      .....
COILS    .....
lnhm-    HHHHTTTTTTTT

SEQ      GAITMATTTTSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGSLAGNEMINGEDEMEMY
SEG      .....xxxxxxx
COILS    .....
lnhm-    .....

SEQ      DDYEDDPKSDYSSENEAPAVSAN
SEG      .....
COILS    .....
lnhm-    .....

```

Prosites for DKFZphtes3_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphtes3_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWMQEMReIKaENPNdMhNtEISKMI GEMWKnMsEEEkM +KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISKILGSRWKSMSNQEKG	644
HMM	PYEdMAeeEKqRYMKEMPeYK* PY+++ +++ + +++ +P+YK		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKFZphtes3_17n18

group: intracellular transport and trafficking

DKFZphtes3_17n18 encodes a novel 782 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein contains TONB_DEPENDENT_REC_1 Pattern and ATP_GTP_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp

Poly A stretch at pos. 2806, no polyadenylation signal found

```

1  GTCCTTTTAA  GTCAGTAAAT  TGAACAAAGT  CGGTTATTCG  GCAAGCAGTT
51  CCTATAAAAA  ACTACATGGC  TAAGTTCTT  AATGATTGAC  CACAAGCAGA
101 TCTTTCACCC  TCGGATCTCT  AGCTACAAAA  GGTCCCCACA  CTGAAGAAGC
151 CACTACCTCC  ACCACCACCA  GCACCACCAC  GTCCAGTGCT  GCTGGCAACC
201 ACTGGGGCAG  CCAAGCGCTC  CACCCTCTCT  CCCACCATGG  CCCGTCAGGT
251 GCGCACCAC  CAGGAGACCC  TGAACAGGTT  TCAGCAGCAG  TCCATCCACC
301 TGCTGACGGA  GCTCCTCAGA  CTGAAGATGA  AGGCCATGGT  GGAGTCTATG
351 TCGGTGGGTG  CCAACCCCTT  GGACATCACC  AGGCGCTTGT  TGGAGGCCAG
401 CCAGCTCCTC  CACCTCAATG  CCAAGGAGAT  GGCCTTCAAC  TGCCTGATCA
451 GCACAGCCGG  GAGAAGTGGC  TACAGCAGCG  GACAGTTGTG  GAAAGAGTCC
501 CTCGCAAAAC  TGTCCGCCAT  TGGGGTGAAC  TCGCCTTACC  AGCTGATCTA
551 CCACTCTTCC  ACAGCCTGTC  TGAGCTTTTC  TCTCTCTGCT  GGAAAAGAAG
601 CCAAGAAGAA  AATAGGCAAA  TCTAGAATA  CAGAAGATGT  CAGCATGCCG
651 CCCCTGCATC  GAGGAGTGGG  AACCCCTGCC  AACAGCCTGG  AGTTCAGCGA
701 CCCCTGCCCT  GAGGCCCGGG  AGAAGCTGCA  GGAGTTGTGT  CGCCACATAG
751 AAGCTGAAAG  GGCCACATGG  AAAGGGAGGA  ATATCTCCTA  CCCCATGATC
801 TTACGAAACT  ACAAGGCAAA  GATGCCCTCT  CATCTAATGT  TGGCCCGCAA
851 AGGAGACTCT  CAGACCCCGG  GTTTACATTA  CCCTCCCACT  GCAGGTGCTC
901 AGACTCTCAG  CCCACCTCT  CACCCATCTT  CTGCCAACCA  TCATTTCAGT
951 CAGCATTGTC  AAGAGGGGAA  GGCACCCAAG  AAGGCCTTCA  AGTTTCATTA
1001 CACCTTCTAT  GATGGCTCCT  CCTTCGTTTA  CTATCCCTCT  GGAACGTCG
1051 CTGTATGTGA  GATCCCCACA  TGCTGCAGAG  GGAGAACCAT  CACCTGCCTC
1101 TTTAATGACA  TACCTGGATT  CTCCTTGCTG  GCCCTATTCA  ATACTGAAGG
1151 CCAGGGCTGT  GTTCACTACA  ACCTAAAAAC  CAGTTGCCCA  TATGTCTTAA
1201 TCTTGGATGA  GGAAGGTGGG  ACCACCAATG  ACCAGCAGGG  CTATGTAGTC
1251 CACAAGTGGA  GCTGGAATTC  CAGGACAGAG  ACCCTGCTTT  CCCTGGAATA
1301 CAAGGTGAAT  GAGGAAATGA  AACTAAAGGT  ACTGGGACAG  GACTCCATCA
1351 CAGTCACCTT  CACCTCCCTG  AATGAGACAG  TAACACTCAC  TGTGTCGGCC
1401 AACAAATTGC  CCCATGGAAT  GGCATATGAC  AAACGGCTGA  ACCGCAGAAT
1451 CAGCAACATG  GACGACAAGG  TGTATAAGAT  GAGCCGAGCC  CTGGCTGAGA
1501 TCAAGAAGCG  GTTTCAGAAG  ACAGTGACTC  AGTTCATTAA  TTCTATCTTG
1551 CTGGCCCGCAG  GTCTGTTTAC  CATTGAATAT  CCCACCAAAA  AGGAGGAGGA
1601 AGAATTTGTT  CGGTTCAAGA  TGAGATCCAG  AACTCATCCC  GAGCGGCTCC
1651 CCAAGCTAAG  TTTATACTCA  GGAGAAAGTC  TTTTACGATC  TCAGTCAGGC
1701 CACCTGGAAT  CCTCAATTGC  AGAGACTTTG  AAGGATGAGC  CTGAGTCTGC
1751 TCCTGTGAGC  CCAGTTCGGA  AGACCACCAA  AATCCACACC  AAAGCCAAGG
1801 TCACATCCAG  AGGGAAGGCC  CGCGAGGGGC  GCAGCCCCAC  CAGGTGGGCG
1851 GCCTTGCCCT  CAGACTGCCC  GCTGGTGCTG  CGGAAGCTCA  TGCTCAAGGA
1901 AGACACCCGT  GCTGGCTGCA  AGTGCCTGGT  GAAGGCGCCC  CTGGTCTCTG
1951 ACGTGGAGCT  GGAGCGCTTC  CTGTTGGCGC  CCCGAGACCC  CAGCCAAGTG
2001 CTGGTGTGTT  GGATCATCTC  AAGCCAGAAC  TACACCAGCA  CTGGGCAGCT
2051 CCAGTGGCTG  CTGAACACTC  TCTACAACCA  CCAGCAGCGG  GGCCGTGGCT
2101 CCCCTGCAAT  CCAAGTCCCG  TATGACTCCT  ACCGCTGCTG  GCAGTATGAC
2151 CTGGACAGCC  CCCTGCAGGA  GGACCCTCCC  CTGATGGTGA  AGAAGAACTC
2201 TGTGGTGCAG  GGGATGATTC  TGATGTTTGC  CGGGGGGAAG  CTCATTTTTG
2251 GGGGCGGTGT  TTTGAATGGA  TATGGCCTCA  GCAAGCAGAA  TCTGCTGAAA
2301 CAGATCTTCC  GGTCTCAACA  GGATTACAAG  ATGGGCTACT  TCCTGCCGGA
2351 TGACTACAAA  TTCAGTGTTT  CCAACTCTGT  CCTGAGCCTG  GAGGATTCTG
2401 AATCAGTCAA  GAAAGCCGAG  TCAGAAGATA  TCCAAGGAAG  CAGCTCCTCA

```



```

2451 TTGGCCCTGG AAGACTATGT GGAQAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACCTAGTTG GAAGAAGCAG GCCTCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGCCCGGGGT GCTGGGGCTT CTGCGCAGCC
2651 CAGCCCTGCC TCCCCGGTCT CCCACCCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGGC TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG
2851 CCG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

```

ORF from 237 bp to 2582 bp; peptide length: 782
Category: putative protein
Prosites motifs: ATP_GTP_A (122-130)
TONB_DEPENDENT_REC_1 (1-44)

```

```

1 MARQVRTHQE TLNRFQQQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSGYS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSA CLSFSLSAGK EAKKIGKSR TTEDVSMPP LHRGVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTAG AQTLSPTSHP SSANHHFSQH CQEGKAPKKA
251 FKPHYTFYDG SSFVYPSGN VAVCQIPTCC RGRITCLFN DIPGFSLLAL
301 FNTEGGQCVH YNLKTSCTPV LILDEEGTT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLKVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRISNMDD KVKMSRALA EIKKRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEFVRF KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTKIHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLRK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPSQVLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCIQCRYD SYRLLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGY LSKQNLKQI FRSQDYKMG
701 YFLPDDYKFS VPNSVLSLED SESVKAASE DIQSSSSSLA LEDYVEKELS
751 LEAEKTREPE VELHPLSRDS KITSWKQAS KK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17n18, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n18, frame 3

Report for DKFZphtes3_17n18.3

```

[LENGTH] 782
[MW] 88030.16
[PI] 9.22
[BLOCKS] BL00286 Squash family of serine protease inhibitors proteins
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 4
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 4
[KW] Alpha_Beta

```


[illegible]

Prosites for DKFZphtes3_17n18.3

PS000001	91->95	ASN_GLYCOSYLATION	PDOC00001
PS000001	182->186	ASN_GLYCOSYLATION	PDOC00001
PS000001	379->383	ASN_GLYCOSYLATION	PDOC00001
PS000001	598->602	ASN_GLYCOSYLATION	PDOC00001
PS000004	403->407	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	511->515	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	652->656	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS000005	177->180	PKC_PHOSPHO_SITE	PDOC00005
PS000005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS000005	450->453	PKC_PHOSPHO_SITE	PDOC00005
PS000005	497->500	PKC_PHOSPHO_SITE	PDOC00005
PS000005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS000005	523->526	PKC_PHOSPHO_SITE	PDOC00005
PS000005	631->634	PKC_PHOSPHO_SITE	PDOC00005
PS000005	723->726	PKC_PHOSPHO_SITE	PDOC00005
PS000005	774->777	PKC_PHOSPHO_SITE	PDOC00005
PS000006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS000006	131->135	CK2_PHOSPHO_SITE	PDOC00006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC00006
PS000006	329->333	CK2_PHOSPHO_SITE	PDOC00006
PS000006	345->349	CK2_PHOSPHO_SITE	PDOC00006
PS000006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS000006	406->410	CK2_PHOSPHO_SITE	PDOC00006
PS000006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS000006	466->470	CK2_PHOSPHO_SITE	PDOC00006
PS000006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS000006	571->575	CK2_PHOSPHO_SITE	PDOC00006
PS000006	693->697	CK2_PHOSPHO_SITE	PDOC00006
PS000006	717->721	CK2_PHOSPHO_SITE	PDOC00006
PS000008	145->151	MYRISTYL	PDOC00008
PS000008	327->333	MYRISTYL	PDOC00008
PS000008	592->598	MYRISTYL	PDOC00008
PS000008	734->740	MYRISTYL	PDOC00008

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphtes3_17n18.3)

DKFZphtes3_18f3

group: testes derived

DKFZphtes3_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```

1  GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51  CCCGACGCGC TCGCGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAAAC TGGCCGCGAG CTCGCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTAC
251 CCTGGGGACC TCGCTGCTGG TGTCGGCCGT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTTCTGCAAC
351 TCCCGGGAGC TCGGAGGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCCG TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGCGGGA GGAACGCCTC CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCCTCAT
551 CCCCAGGCGG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCCTGCAC CGGGGCTCTG
651 GACGAACCTCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CACGACCTCA AGATCTCTGC TGACCAGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAATCATATC
801 TCATGGGATG CTCCAGAAAT TGTAAGTCCC TTAGGAAAAC ACCAAGCTGG
851 GTTAGGAGCC GAAGGCAAAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCG
901 GTCCCCAAAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCAAAAACT
1001 CCTTTTCTTT TATCAAAAAC TTTCTGTCTA AACACAGCTG GGCAGGCACT
1051 CCTGTTTTAA AGTTATTTTC GGGTCCCTGA CCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TCGGTGAGGA AGGCATTTGC CTCTATTCCA
1201 GAATTTCTGA TACAAAGAAC TCCAGAATCC AGAGCAATC AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCCAGA GAGGACCCCT GTGCTGATAT
1301 CTCTCTCTCT TCCCTTTCCC CTCAGCTTAC TTACTCCAG ATGCGGCTG
1351 GGTATGAAGT AGGCCTTTCC TGAGTGGCTC CCAATCCAGT CTCTCAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCATC TAAGTCTCTG TCCCTCACAT
1451 GAAGCTGAGG GCCAGATAGA TGGAGCGACT GCCAATTCA TTTCCGACA
1501 TCATTGTGTT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCTGGGC
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCCTGTT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CCTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTGATA AAGGCAAAAT
1751 CTTGGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTTCTCT GGGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA AATCACTACA ACAAATGCTA ACTTCTAATC CCCTTGATGA
1901 GCTTTCACGA AGTCTCACGG CTTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTCACAGA TGACCAAGGA CAGACTGTGT CCCAGAAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCACGCGTA CGTGCACCCT GGGGCAGTGT CTCACCGTAT
2051 GAATAAGGGA TGTAACACTA AAAGCCCAT AGGGGGCAGT GTTCCCGCC
2101 TGTGTAGAAA ACTGTACAG AAAGGATCCT ATATGAAGTT CCTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAAA GTGCCAGTCC ATGTATTTT
2201 TATTTATTTT AAGTTGTGTA TTAAATTTT AATTATTGTT TAGTGTGTTG
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTCTCTA GTGTTTAAAG TGATCCCTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAAGATAA AAAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAACTTGCA AGCCTGATGT CCTATCAAGT TATGTCTTCT GGGTGACAGA
2701 CAAAATAGCT TGTCTTATGG TGGTGATGTG TTGCATTTTC ACTTTGGGGT

```



```

2751 CTGTAAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTC
2801 TTAACAACCTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCCCTCTG TTACTTAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCCT TCCCAGTGAT GGGGCAGTAT GTCTGAGGAA
2951 GTATAATTTT AGTACTGGGG TCGGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTT TATAAATTGC AATTGGTCTG TATGCTGGTT TATTTTAAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTC ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTGTGAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCAGTGCAA CCTCCGACTC
3201 CTGGGTTCAA GCAATTCCTC TGCCTCAGCC TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTTGGCCA GGCTGGTCTT GAATCCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGCTGG GATTACAGGC ATGAGCCACC
3401 TCGCCTGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATTGTTTTGT TTCACAATCA TTTTAAATCA TTTTGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCTTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATCTT TCCAATGCCT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACCTTTCC TCTCATATTT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATTG CAGTGGCCCA CTTTATTCTT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAATCA AGAAATACAT TTGAATAATT
3801 ATAATTAATC GTTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTCAG TTAGATTGAT TCTAGAAACA AATATTTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTTGAAA TGGTATGTTG TAAAAGAGAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTCTTAAAT GTTTAAGTGC TTCTCTGTTA GGTTCGCGG CTTGCAATCA
4151 TTTGAATTGT TCTGTTTCAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATTCA ATAGAAGTGC TCCATTAAAA AAATAATCCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCAGTGGC CTCGGGCACC TGTCATTTCC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAAGAGAG
4351 GAGTGGTGTG AGAAGTCTCC CTAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTCA ATATTTTAGT GTGAATTGAG ACACTGAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTTATT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

BLAST Results

Entry HSG27587 from database EMBL:
human STS SHGC-32548.
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:
human STS EST303564.
Score = 1417, P = 8.7e-58, identities = 285/287

Medline entries

No Medline entry

Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194
Category: questionable ORF
Classification: no clue

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 2

PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PA1 1 gene: "COL1A1"; Human proalpha 1 (I) chain of type I procollagen mRNA (partial)., N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)
Length = 779

HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAARAAALPGTAA--GPPRPAAPGA--APARGGPAPGAPALPRSQGR 62
G+ G PG + AR PG GPP PA P GA AP G A A P SQ
Sbjct: 230 GDLGAPGPSGARGERGFPGERGVEGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQAGP 289

Query: 63 QLAERNRPRRRHAGALAQPHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
L G P RGA PG GD +GA G + G VR L + PG A
Sbjct: 290 GL---QGMFGE-RGAAGLPGPKGDRGDAGPKGADGAPKDG----VRGLTGPIGPPGPAG 341

Query: 123 GAGDRGHL-P-GP-----DARPELPRVFLPLAGLRGPPAA 156
GD+G P GP D +P P P AG GPP A
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPGAAPARG-----GPAPGAPALPRSQRG 61
G G PGAA R P AGPP P P G ++G GPA G P + P G
Sbjct: 434 GATGFPGAA-GRVGPFGPSGNAGPPGPPGPAGKEGSKGRGETGPA-GRPGEVGPFGPPG 491

Query: 62 RQLAERNRPRRRHAGALAQPHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 121
A G P G PG PG RG G +RG R L PG +
Sbjct: 492 P--AGEKAPGAD-GPAGAPGTGPGQGIAGQRGVVGLPGQORGE----RGFPGL---PGPS 541

Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160
G+G R P P + GL GPP + RE
Sbjct: 542 GEPGKQGPSGASGERGPPG---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPALPRSQGR-R 62
G G PG AR +A PG A G P A PPG + GP PG P A +G R
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATGFPGAAGRVGPPGPS-GNAGP-PGPPGPAGKEGSKGPR 472

Query: 63 QLAERNRPRRRHAGALAQPHPGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120
GRP G + PG PG GA G G + ++ LPG
Sbjct: 473 GETGPAGRP---GEVGPFGPPGPAGEKGAPGADGPAGAPGTGPGQGIAGQRGVVGLPGQ 528

Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154
G+RG LPGP + P +G RGFP
Sbjct: 529 R---GERGFPGLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAARAAALPGT---AAGPPRPAAPGAAPARG--GPA--PGAPALPRSQR 60
G G PG + PG A+GP P PPG G G A PG P + P +
Sbjct: 29 GPPGAPGPGQFGQPPGEPGEGASGPMGPRGPPGPKNGDDGEAGKPRGGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHAGALAQPHPGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115
G R L G P + HRG G GD +G G G + R L
Sbjct: 89 GARGLPCTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGEPGSPGENGAPGQMGPRGLPGFP 148

Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
GAA G AG+RG +PGP P AG +GPP A
Sbjct: 149 GPKGAAGEPGKAGERG-VPGPPGAVG--PAGKDGEAGAQGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAARAAALPGTA-----AGPPRPAAP---PGAAPARGGPAP-GAPALPR 57
G AG PGA A PG A AGPP PA P PG G P P GA A P
Sbjct: 374 GFAGPPGADGQFGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVAGPAGPKARGSAGPP 433

Query: 58 SQRGQLAERNRPRRRHAGALAQPHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G A P G PG PG +G G GR V
Sbjct: 434 GATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGSKGRGETGPAGRPGEVGP----- 486

Query: 118 PGAAEGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152
PG AG++G PG D A P P +AG RG
Sbjct: 487 PGPPGPAGEKG-APGADGPAGAPGTGPG-QGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPP--AAPAR-GGPAP-GAPAQALPRSQRGR 62
 GE G G A + LPG A GPP A PG P G P P GA + +RG
 Sbjet: 194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEQGVPGDLGAPGSPGARGERGFPGERGV 252

Query: 63 QLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAE 122
 + PR GA G GD A G+ G +G R A L PG
 Sbjet: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQCAPGLQMPGE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157
 GDRG GP D P V L G GPP A
 Sbjet: 308 --GDRGDA-GPKGADGAPGKGV-RGLTGPPIGPPGA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03
 Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPAQALPRSQRG 61
 NG+ GEAG PG R P A G P A PG RG GA A P +G
 Sbjet: 67 NGDDGEAGKPGRP-GERGPPGPQARGLPAGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKG 125

Query: 62 RQLAE-RNGRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSL-----ADLL 115
 + NG P + G PG PG A G G G V A
 Sbjet: 126 EPGSPGENGAPGQ-MGPRGLPGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQ 184

Query: 116 QLPGAAGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
 PG A AG+RG GP A P F L G GPP A
 Sbjet: 185 GPPGPAGPAGERGE-QGP-AGSPG----FQGLPGPAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNNGEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
 E GE G PG R LPG GP A PG A RG P P GA A +
 Sbjet: 126 EPGSPGENGAPGQMGPR---GLPGFP-GPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEA 181

Query: 61 GRQLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
 G Q P RG G PG G+ G G G+ DL PG
 Sbjet: 182 GAQGPFPAGPAGERGEQGPAGSPG--FQGLP-GPAGPPGEAGKPGEQGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132
 + G+RG PG
 Sbjet: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQLAE 66
 GEAG G A R A PG G P P P G A GP PGA Q + + G A+
 Sbjet: 347 GEAGPSGPAGTRGA---PGDR-GEPPPGPAGFA----GP-PGADGQPGAKGEPGDAGAK 397

Query: 67 RNRGRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEGAGD 126
 + P G PG G++ A +GA G G + A + PG + AG
 Sbjet: 398 GDAGPPGPAGPAGPPGPIGNVGAAPGPKGARGSGAGPPGATGFPGA-AGRVGPPGPGSGNAGP 456

Query: 127 RGHLPDPDARD 137
 G PGP ++
 Sbjet: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQL 64
 G AG PGA A G GP P P G A ARG P P Q PR +G
 Sbjet: 608 GPPGAPGAPGVPVGPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRBKGZTG 662

Query: 65 AERNRPRRRHRG---ALAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLA-DLLQ-LPG 119
 + + + HRG PG PG GA G RG S D L LPG
 Sbjet: 663 ZZGBRGIKGRGFSGLQGPVPPGSPGEQGPSGASGPAGPRGPPGSAGSPGKDGLNGLPG 722

Query: 120 AAGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQ 168
 G RG GP A P P P G GPP+ L +P Q
 Sbjet: 723 PIGPPGPRGRTGDAGP-AGPFGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02
 Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPA---QALPRSQRGR 62
 G AG PG A R PG A GP A G A A+G P P PA + P G
 Sbjet: 152 GAAGEPGKAGERGVPGPPG-AVGP---AGKDGEAGAQGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
 Q P G + G PGDL A G G RG R + PG A
 Sbjct: 208 QGLPGPAGPPGEAGKPGEQGVPGDLGAP---GPSGARGERGFPE-RGVEGP---PGPAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154
 G G PG D + P G +G P
 Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02
 Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAARAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60
 G G PG + PG A+GP P PPG G G A PG P + P +
 Sbjct: 29 GPPGAPGPQGFQPPGEPGEPGASGPMGPRGPPGKNGDDGAGKPRPGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
 G R L G P + HRG G GD +G G G + L
 Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPGRG-LPGF 147

Query: 118 PGAEGAGDRG 128
 PG AG+ G
 Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02
 Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAARAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQAL 55
 G G PGA R A PG A G P P P G + RG GPA P PA A
 Sbjct: 587 GRDGSPPGAKGDRGETGPAGAPGPPGAPGAPGVPAGKSGDRGETGPAGPIGPVGPAGAR 646

Query: 56 -----PRSQRQLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHV 108
 PR +G + + + HRG G PG + +G G G
 Sbjct: 647 GPAGPQGPBGKGTGZZGBRGIKGHRGFSGLQGPPGPPGSPGEGQPSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAEGAGDRG--HLPDPDARDPELPRVFLPLAGLRGPP 154
 PG+A G G LPGP P PR AG GPP
 Sbjct: 706 -----PGSAGSPGKDGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02
 Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAARAAALPGTAA----AGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
 G G G R AA LPG AGP PG RG P G P A +
 Sbjct: 287 GAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 346

Query: 61 GRLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
 G A +G P RGA +PG PG GA G +G + D
 Sbjct: 347 GE--AGPSG-PAGTRGAPGDRGEPGPPGPFAGFAGPPGADGQPGAKGEPGDAGAKGDAGP- 402

Query: 118 PGAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159
 PG A AG G + A P+ R G G P AA R
 Sbjct: 403 PGFAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02
 Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPA-QALPRSQRGR 62
 +G G PGA + PG G PA PG A G P P PA ++ R + G
 Sbjct: 574 SGREGAPGAEGSPGRDGSPPGAKGDRGETGPAGAPGPPGAPGAPGVPAGKSGDRGETGP 633

Query: 63 QLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
 P RG G G+ +G G RG H R + L PG
 Sbjct: 634 AGPIGPVGPAGARGPAGPQGRGB-----KGZTGZZGBRGIKGH-RGFSGLQGPPGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
 G++G P A P AG RGPP +A
 Sbjct: 687 SPGEQG--PS-GASGP-----AGPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02
 Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPFRPAAPPGAAPARGGPA-PGAPAQALPRSQRQLAERNRPRRRHRLAQA 80
 P G P P PG +G P PG P + P RG G P ++ G +
 Sbjct: 21 PSGPRGLPGPPGAPGPQGFQPPGEPGEPGASGPMGPRGPP-----GPPKNGDDGEAGK 75

Query: 81 PGHPGDLAA-GV--GRGAGGGHSRRGRHHVRSADLLQLPGAEGAGDRG--LPGPDA 135
 PG PG+ G RG G G H R + L G A AG +G PG +
 Sbjct: 76 PGRPGERGPPGPQARGLPGTAGLPGMKGH-RGFSGLDGAKGDAGPAGPKGEPGSPGENG 134

Query: 136 RDPDL-PRVFLPLAGLRGPPAAA 157
 ++ PR LP G GP AA

Sbjct: 135 APGQMGPARG-LP--GFPKPGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPALPRSQGRQLA 65
GEAG G A R A G GPP PA G A G P A G P A + G
Sbjct: 347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405

Query: 66 ERNGRPRRRHRLAQPDPDLAAGVGRGAGGGHSRRGR--HHVRSADLLQLPGAA-- 121
P G + PG G + GA G GR A PG A
Sbjct: 406 AGPAGPPGPIGNVGAAPGKARGSGAGPPGATGFPGAAGRVGPPGSGNAGPPGPPGPAGK 465

Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156
EG+ G RG GP R E+ P AG +G P A
Sbjct: 466 EGSKGPRGET-GPAGRPGEVGPFPGPAGKAGPA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARGGPAPGAPAL-PRSQ 60
G G PGA R A PG A G P P P G + RG P P + P R
Sbjct: 587 GRDGSPPGAKGDRGETGPAGAPGPPGAPGAPGVGPKSGDRGETGPAGPIGVGPAGAR 646

Query: 61 GRQLAERNRPRRRHRLAQPDPDLAAGVGRGAGGGHSRRGRH--HHVRSADLL 115
G A G PR +G + G G +G G G A
Sbjct: 647 GP--AGPQG-PRGBKGTGZZGBRGIKHGRGFSGLQGPPGPPGSGEQGSGASGPAGPR 703

Query: 116 QLPGAAGAGDRG--HLPDPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LPGP P PR AG GPP
Sbjct: 704 GPPGSAGSPGKDLNGLPGPIG--PPGPRGTGDAGPAGPP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQGRQ-LA 65
G G PG A + A G A P P P G A RG G P Q R +RG L
Sbjct: 485 GPPGPPGAPGKAGADGPAGAPGTGP-PQGIAGQGR--VVGLPGQ---RGERGFPLP 538

Query: 66 ERNGRPRRRH--RGALAQPDPDLAAGVGRGAGGGHSRRGRHHVRSADL 114
+G P + GA + G PG + AG GR GA G GR + D
Sbjct: 539 GPSGEPGKQGPSGASGERGPPGMPGPPGLAGPPGESGREGAPGAEGSPGRDGSFGAKGDR 598

Query: 115 LQL-PGAAGAGDRGHLPGP 133
+ P A G PGP
Sbjct: 599 GETGPAGAPGPPGAPGAPG 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG--GPAP--GAPALPRSQ 60
G+AG GA A + G GPP PA PG G GPA GAP R +
Sbjct: 311 GDAGPKGADGAPGKDVRLTGPPIGPPGAPGDKGEAGPSGPAGTRGAPGD---RGEP 367

Query: 61 GRQLAERNRPRRRHRLAQPDPDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 120
P G G PGD A G G G + ++ PG
Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPIGNV----APGP 423

Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAVRE 160
G G PG RV P AG GPP A +E
Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGSGNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQGRQLAE 66
G+AG PGA ++ A L G G A PG RG P A P R L
Sbjct: 275 GDAGAPGAPGSQAGPLQGMP-GERGAAGLPKPKGDRGDAGPKG-ADGAPGKDVRLTG 332

Query: 67 RNRGPRRRHRLAQPDPDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAAGAGD 126
G P G PG G+ G G RG A PGA G
Sbjct: 333 PIGPP----GPAGAPGDKGEAGPSGPAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154
+G PG A+ P P AG GPP
Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

ORF from 12 bp to 755 bp; peptide length: 248
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (17-39)
 LEUCINE_ZIPPER (24-46)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 3

TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802_1-gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.
 Length = 331

HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVRLRLREVARRLERLRRRSLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLLVSA 89
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A
 Sbjct: 91 KIQESIEKLRLALANGIEEVHRGCTISNVVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150

Query: 90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIAATCQDQMR 132
 G+G+ A IT+ + + +S E + AT D+++
 Sbjct: 151 AGVGLGAASAVTGITTSIVEHSYTSSAEAE-ASRLTATSIDRLK 193

Pedant information for DKFZphtes3_18f3, frame 2

Report for DKFZphtes3_18f3.2

[LENGTH] 193
 [MW] 19708.24
 [pI] 11.90
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQR
 SEGxx
 PRD cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccchhhhhh

SEQ GRQLAERNRPRRRRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA
 SEGxx
 PRD hhhhhhccchhhhhhhhhcccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQFCLLHRLWLWTW
 SEGxx
 PRD cchhhhhhhhhccccchhhhhhhhhc

SEQ LPHPQAGGGGHQ
 SEG xxxxxxxxxxxxxxx
 PRD ccccccccccccc

(No Prosite data available for DKFZphtes3_18f3.2)

(No Pfam data available for DKFZphtes3_18f3.2)

Pedant information for DKFZphtes3_18f3, frame 3

Report for DKFZphtes3_18f3.3

```

[LENGTH]      248
[MW]           27162.56
[pI]           9.92
[PROSITE]      LEUCINE_ZIPPER 2
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY  30.65 %
[KW]           COILED_COIL     12.10 %

SEQ  MGMRPAAREPHGPDALRRFQGLLLDRRGRRLHRQVLRRLREVARRLRLRRSLVANVAGS
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...XXX
PRD  cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
COILS .....
MEM  .....

SEQ  SLSATGALAAIVGLSLSPVTLGTSLLSAVGLGVATAGGAVTITSDSLIFCNSRELRRV
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...
PRD  cchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhhhhhh
COILS .....
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ  QEIAATCQDQMREILSCLEFFCRWQCGDRQLLQCGRNASIALYNSVYFIVFFGSRGFLI
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc
COILS .....
MEM  .....

SEQ  PRAEGDTKVSQAVLKAQKIQKLAESLESCTGALDELSEQLSRVQLCTKSSRGHDLKISA
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
MEM  .....

SEQ  DQRAGLFF
SEG  .....
PRD  hhhhhccc
COILS .....
MEM  .....

```

Prosites for DKFZphtes3_18f3.3

```

PS00029      17->39  LEUCINE_ZIPPER      PDOC00029
PS00029      24->46  LEUCINE_ZIPPER      PDOC00029

```

(No Pfam data available for DKFZphtes3_18f3.3)

DKFZphtes3_1817

group: cell structure and motility

DKFZphtes3_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```

1  GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCAGCAG CGGACGAGGG
51 AGGTGCCCCG GTCGCCCCAG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
101 TAAGAGATTT GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
151 AGACCTCCTG AAAAATCCTT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
201 ACTTGTGCAG CAAAGTGGCC CAAATCCATG GCATTGTCTT AGTACCTGTC
251 AAAGGAAGCC TGTCGAGCAG CATCCAGTCT ACTTGTCAGT TTGAGTCCTA
301 CATTTTGATA CCTGTGGAAG AGCATTTTCA GACCTTAAAT GGAAAGGATG
351 TCTTTATTCA AGGGAACAGG ATTAAATTAG GAGCTGGTTT TGCCTGTCTT
401 CTCTCAGTGC CCATTCTCTT TGAAGAACT TTCTACAATG AAAAGAAGA
451 GAGTTTCAGC ATCCTGTGTA TAGCCCATCC TTTGGAAAAG AGAGAGAGTT
501 CAGAAGAGCC TTTGGCACCC TCAGATCCCT TTTCCCTGAA AACCATTGAA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTTG ACAGGAACAT
601 CGCCTCTTTC CATCGAACAT TCCGAGAATG CGAGAGAAAG AGCCTCCGTC
651 ACCACATAGA CTCAGCGAAT GCTCTCTACA CCAATGCGCT CCAGCAGCTT
701 CTGAGGGACT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
751 CCTGATGAAG CAGGCAGTGG AGATATACGT CCATCATGAA ATTTACAACC
801 TGATCTTTAA ATACGTGGGG ACCATGGAGG CAAGTGAGGA TGCGGCCCTT
851 AACAAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAAGA GAGCTGGCTC
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTGCGAAAA
1001 GTGGTGCAGC TCATTACACA GTCTCCAAGC CAGAGAGTGA ACCTGGAGAC
1051 CATGTGTGCT GATGATCTGC TATCAGTCTT GTTATACTTG CTGTGAAAAA
1101 CGGAGATCCC TAATTGGATG GCAAATTTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGGCAAAGGA TGAAGTGGGA TACTGCCTGA CCTCATTGCA
1201 AGCTGCCATT GAATATATTG GGCAAGGAAG CCTCTCTGCT AAACCCCTG
1251 AGTCTGAGGG ATTTGGAGAC AGGCTGTTCC TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCCACCGAC TGCCCTGTTA AGCACATTGC
1351 ATCAGGTAAC CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATCA CGTCCAAAAG ATGTGTCACC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTGAAT GATCCCTCAG TTGTCACTCC
1501 ATTCTCCAGA GACGACAGGG GGCACACCCC TCTCCATGTG GCTGCTGTCT
1551 GTGGGCAGGC ATCCCTCATC GACCTCCTGG TTTCCAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCCACCTGG CCTGTCAGAA
1651 GGGCTACCAG AGCGTGACGC TGCTGTGTGT GCACTACAAG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACCTGGC CTGCACCTAC
1751 GGCCACGAGG ACTGTGTGAA GGCTCTGGTT TACTACGACG TGGAGTCGTG
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGCTGCA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCCCTA AGTGTGCATT
1951 AAACCTCAAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCCTGTGC AGTCCCCGCA GCGCTCCGTG
2051 GACTCCATCA GCCAAGAGTC CTCCACTTCC AGCTTCTCCT CCATGTCAGC
2101 CGGCTCAAGG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGCTGAT GGAGATCTAG AAATGGTGCG TTACCTGTTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGCAGC
2251 AGACCCCGAA TTCTGTCACC CGTTGTGCCA GTGCCCCAAG TGTGCCCCAG
2301 CTCAGAGAGG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCCTC CCCGCTGCAT GTCGCGCCCG TGCACGGCCG
2401 GGCGGACCTC ATCCGCTTCC TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCCTGCCA GCAGGGCCAC
2501 TTTCAAGTGG TGAAGTGCTT GTTAGATTCT AATGCAAAAC CCAATAAGAA
2551 GGACCTTCAG GGAACACGCG CCCTCATTTA CGCCTGCTCC GGTGGCCATC
2601 ACGAGCTTGT GGCAGTGTCT CTACAGCAGC GGGCCTCCAT TAACGCTTCT
2651 AACAATAAGG GCAACACAGC GCTGCACGAG GCTGTGATTG AAAAGCACGT

```



```

2701 CTTCTGGGTA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTCAC TGTAAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGGT TCTTTGCATG
3051 AGCCAGGGAG GCAAAGTGTC ACACTGAGAC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACGGATGCTG CGGAGACACA
3201 CGGTAGAGGA TGCGGTCGTG TCCCAGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTTCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTCAGCACC AAGTTCTCTG AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCCT CTTCAAAGCT AATGAATACA
3501 ATTGAACACG ACAAATAATC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAAC CACAGCTAGA ATTCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAATGTGTA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCATT TTGATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGTCAC CTGTAGCAGT ATTTCTAGAA ACATCCCTTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTGTG TCAAATTCAA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTC CTTTTTATAA ATTTCTATAG ATTATACTGT TATTTTATG
4201 TTATTGGCTC AGAGCTACAC GTATATGGGT TTGTCTGAG TCCGTTTCA
4251 AATGACCTTG TGATAGGGAA ATGGTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTTCTACAA AGTAATTTAT
4351 GATTTCTAAT TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAG
4401 AAACAGTAAA CTTTATGATT CTAAAAAAA AAAAAAAA AAAAAAAA
4451 AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA
4501 G

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050
 Category: similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (945-953)

```

1  MALYDEDLLK NPFYLAQKC RPDLCCKVAQ IHGIVLVPCCK GSLSSSIQST
51  CQFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILFEETF
101 YNEKEESFSI LCIAHPLEKR ESSEELAPS DPFSLKTIED VREFLGRHSE
151 RFDNRNIASFH RTFRECEKRS LRHHIDSANA LYTKCLQQLL RDSHLKMLAK
201 QEAQMNLMKQ AVEIYVHHEI YNLIFKYVGT MEASEDAAFN KITRSLQDLQ
251 QKDIGVKPEF SFNIPRAKRE LAQLNKCTSP QOKLVCLRKV VQLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWMA NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIRQGSLSAK PPESEGFGRD LFLKQRMSSL SQMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFDDC EKLVSGRIND
451 PSVVTPEFRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLAQCQGYQS VTLLLLHYKA SAEVQDNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLDI GNEKGDTPH IAARWGYQGV IETLLQNGAS TEIQNRLKET
601 PLKCALNSKI LSVMEAYHLS FERROKSSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSAADPEF CHPLCQCPKC APAQKRLAKV PASGLGVNVT SQDGSSPLHV
751 AALHGRADLI RLLKKGANA GARNADOAVP LHLACQQGHF QVVKCLLDSN
801 AKPNKKDLGS NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFFVE LLLHGGASVQ VLNKRQRTAV DCAEQNSKIM ELLQVVPSCV

```


901 ASLDDVAETD RKEYVTVKIR KKWNKLYDL PDEPFTROFY FVHSAGQFKG
 951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRQNNLPAQS GSHAAEKGNS
 1001 DWPERPGLTQ TGPGRHRLR RHTVEDAVVS QGPEAAGPLS TPQEVASRS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_1817, frame 2

TREMBL:HSU43965_1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin
 G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
 ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE
 ANKYRIN), N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =
 380, P = 8.2e-31

>TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
 ankyrin (variant 2.1)
 Length = 1,719

HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31
 Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKAS 521
 +G+T LH+AA+ GQ ++ LV+ GA VNA G TPL+AA Q+ + V LL A+
 Sbjct: 77 KGNTALHIAALAGQDEVRELVNYGANVNAQSQKGFPLYMAAQENHLEVVKFLENGAN 136

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYYDVES-CRL----- 558
 V +G TPL +A GHE+ V L+ Y + RL
 Sbjct: 137 QNVATEDGFTPLAVALQQGHENVVAHLINYGTKGKVRPLALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615
 D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+
 Sbjct: 197 PNPVLSKTGFTPLHIAAHYENLNVAQLLLNRGASVNFPTQNGITPLHIA--SRRGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673
 L +R + E + + ++ S + G+ Q +TK +
 Sbjct: 255 V-RLLLDRGAQI-ETKTKDELTPHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPA 732
 A GD L+ VR LL++ E ++D T+ P H C R+AKV
 Sbjct: 312 ---AAQGDHLDVRLLLQYDAE-IDDI--TLDHLTP--LHVAHC-----GHRVAKVLL 358

Query: 733 S-GLGVNVTSQDGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQGGHFQ 791
 G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH
 Sbjct: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMELLKGTASIDAVTESGLTPLHVASFMGHL 418

Query: 792 VVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAV 851
 +VK LL A PN ++ TPL A GH E+ LLQ+ A +NA T LH A
 Sbjct: 419 IVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPLHCAA 478

Query: 852 IEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQV 896
 H +V+LLL + A+ + T + A + + +L ++
 Sbjct: 479 RIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30
 Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKASAEV 524
 TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +
 Sbjct: 274 TPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYYDVE-----SCR----- 557
 + TPLH+A GH K L+ + +C+
 Sbjct: 334 ITLDHLTPLHVAACHGHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMELLK 393

Query: 558 ---LDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSV 614
 +D E G TPLH+A+ G+ +++ LLQ GAS + N ETPL A + V
 Sbjct: 394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLL 674
 + Y L + + + Q+P I + +A T L
 Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHNMMVKLLLENNANPNLATTAGH----TPLH 508

Query: 675 RAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734
 A +G +E V LLE ++ A T P H + K A+ L +
 Sbjct: 509 IAAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAKYGKVRVAELLER----D 559

Query: 735 LGVNVTSQDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVVK 794
 N ++G +PLHVA H D+++LLL G + + + PLH+A +Q +V +
 Sbjct: 560 AHPNAAAGKNGLTPLHVAVHHNNLDIVKLLPRGGSFSPAWNGYTPHIAAKQNQVEVAR 619

Query: 795 CLLDSNAKPNKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNGNTALHEAVIEK 854
 LL N + + G TPL A GH E+VALL A+ N N G T L H E
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 HV V ++L+ HG V + T + A N K+++ L
 Sbjct: 680 HVPVADVLKHHGVMVDATTRMGYTPHVAASHYGNIKLVKFL 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDVTQKMCPL-CFCDDCEKLVSGRLNDPSVVTFFSRD 460
 HIAS GN V LL + + + PL C + +S L D ++
 Sbjct: 244 HIASRRGNVIMVRLLLDRGAQIETKTKDELTPHCAARNGHVRRISEILLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKA 520
 G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A
 Sbjct: 303 KNGLSPIHMAAQGDHLCVRLLLQYDAEIDDITLDHLTPHVAACHGHHRVAKVLLDKGA 362

Query: 521 SAEVDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGYQGV 580
 + NG TPLH+AC H ++ L+ +D E G TPLH+A+ G+ +
 Sbjct: 363 KPNRSLNGFTPLHIACKNNHVRVMELLK---TGASIDAVTESGLTPHVASFMGHLPI 419

Query: 581 IETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637
 ++ LLQ GAS + N ETPL A ++++ + +K + P+ R
 Sbjct: 420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ----SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTE 693
 ++ + E++ + +AG VE +L + + +T
 Sbjct: 480 IGHNTMMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query: 694 EDEDAEDTVSAAD---PEFCHPLCQ-----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741
 + V A+ HP P A L V G + +
 Sbjct: 540 LHVAKYGKVRVAELLERDAHPNAAAGKNGLTPLHVAVHHNNLDIVKLLPRGGSFSPSPA 599

Query: 742 QDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVVKCLLDSNA 801
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 600 WNGYTPHIAAKQNQVEVARSLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNGNTALHEAVIEKHVFEVVEL 861
 N + SG TPL GH + +L++HG ++A+ G T L H A ++ +V+
 Sbjct: 660 NGNLGNKSGLTPLHLVAQEGHVPVADVLKHHGVMVDATTRMGYTPHVAASHYGNIKLVKF 719

Query: 862 LLLHGASVQVLNK 874
 LL H A V K
 Sbjct: 720 LLQHQAADVNAKTK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDVTQK---MCHPLCFCDDCEKLVSGRLNDPSVVTFFS 458
 H+AS G+ K V LL +E + T +K H +++V +N + V +
 Sbjct: 50 HLASKEGHVKMVEVLLHKEIILETTTKKGNTALHIAALAGQ-DEVVRELNVYGANVN--A 106

Query: 459 RDDRGTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHY 518
 + +G TPL++AA ++ L+ GA N G TPL +A Q+G+++V L++Y
 Sbjct: 107 QSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQGHENVAHLINY 166

Query: 519 KASAEVDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGY 577
 +V+ P LH+A ++D A V + D+ ++ G TPLHIAA +
 Sbjct: 167 GTKGKVR-----LPALHIAAR--NDOTRTAAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQS 634
 V + LL GAS + TPL A N ++ ++ E + K P+
 Sbjct: 219 LNVQALLNLRGASVNETPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDELTPHLC 278

Query: 635 PQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMVRYLLEWTE 693
 R+ E + + A +TK + A GD L+ VR LL++
 Sbjct: 279 AARNGHVRISEILLDHGAPIQA-----KTKNGLSPIHM----AAQGDHLCVRLLLQYDA 329

Query: 694 EDLEDAE-DTUSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729
 E ++D D ++ C H ++ P C R + +
 Sbjct: 330 E-IDDITLDHLTPLHVAHCHGHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDQGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGG 788
 + +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G
 Sbjct: 389 LLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAG 448

Query: 789 HFQVVKCLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNKGNTALH 848
 H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH
 Sbjct: 449 HTEVAKYLLQNAKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGHTPLH 508

Query: 849 EAVIEKHVFVVELLLHVASVQVLNKRQRTAVDCAEQNSKIM--ELL 893
 A E HV V LL AS + K+ T + A + K+ ELL
 Sbjct: 509 IAAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPFSVTPFSRDD 461
 H+A+ G + E LL ++ H + PL L +L P +P S
 Sbjct: 541 HVAAYGKVRVAELLERDAHPNAAGKNGTLP HVAVHHNNLDIVKLLPRGGSPHSPAW 600

Query: 462 RGHTPLHVAACGQASLIDLVSCKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
 G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+
 Sbjct: 601 NGYTPLHIAAKQNQVEVARSLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 522 AEVDQNNNGTPLHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHIAARWGYQGV 581
 + + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++
 Sbjct: 661 GNLGNKSGTLP HLVAQEGHVPVADVLIKHGVS---MVDATTRMGYTPLHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRKLTPL 602
 + LLQ+ A + +L +PL
 Sbjct: 718 KELLQHQADVNAKTKLGSPL 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDQGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVV 793
 G VN T Q+G +PLH+A+ G ++RLLL GA + D+ PLH A + GH ++
 Sbjct: 229 GASVNFTPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTDELTPHCAARNGHVRS 288

Query: 794 KCLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNKGNTALHEAVIE 853
 + LLD A K +G +P+ A G H + V LLLQ+ A I+ T LH A
 Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDCVRLLLQYDAEIDDITLDHLTPLHVAHC 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909
 H V ++LL GA + + LN + C + + ++MELL AS+D V E+
 Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMELLKKG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14
 Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPFSVTPFSRDD 461
 HIA+ G + V LL +E +K PL K+ L P +
 Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHPNAAG 567

Query: 462 RGHTPLHVAACGQASLIDLVSCKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
 G TPLHVA ++ LL+ +G ++ ++G TPLH+A ++ V LL Y S
 Sbjct: 568 NGTLP HVAVHHNNLDIVKLLPRGGSPHSPAWNGYTPLHIAAKQNQVEVARSLQYGG 627

Query: 522 AEVDQNNNGTPLHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHIAARWGYQGV 581
 A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V
 Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLSKQANG---NLGNKSGTLP HLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPV-QSPQR 637
 + L+++G + R+ TPL A N K++ + + + K +P+ Q+ Q+
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNAKTKLGSPLHQAQQ 744

Query: 638 S-VDSISQ--ESSTSSFSMSAGSRQEETK--DYREVEKLLRAVAD 679
 +D ++ ++ S S G+ K Y V +L+ V D
 Sbjct: 745 GHTDIVTLLKNGASPNVSSDGTPLAIKRLGYISVTDVLKVVT 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDQGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVV 793
 G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V
 Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLSKQANGNLGNKSGTLP HLVAQEGHVPVA 684

Query: 794 KCLDSNAKPNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKNGNTALHEAVIE 853
 L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLLLQH QADVNAKTKLGYSPHLHQAQQ 744

Query: 854 KHVEFVVELLLH GASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896
 H +V LLL +GAS ++ T + A++ + ++L+VV
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTP LAIAKRLGYISVTDV LKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHDKDTVQKMC--PLCFDDC-EKLVSGRLNDPSVVT PFSR 459
 H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR
 Sbjct: 310 HMAAQGDHLDVCVRLLLQYDAEIDDIT-LDHLTPLHVAACHGHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
 G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL
 Sbjct: 368 ALNGFTPLHIACKKNHVRVMELLKLTGASIDAVTESGLTPLHVASFHGLFIVKNLLQRG 427

Query: 520 ASAEVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARWGYQG 579
 AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+
 Sbjct: 428 ASPNVSNVKVETPLHMAARAGHTEVAKYLLQ---NKAKVNAKAKDDQTPHLCAARIGHTN 484

Query: 580 VIETLLQNGASTEIQNRLKETPLKCA 605
 +++ LL+N A+ + TPL A
 Sbjct: 485 MVRLLLENNANPNLATTAGTTPHLIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33
 Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLDSNAK 802
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 601 NGYTPLHIAAKQMQVEVARSLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLLSKQAN 660

Query: 803 PNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKNGNTALHEAVIEKHVFEVVELL 862
 N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+ L
 Sbjct: 661 GNLGKNSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 L H A V K + + A Q ++ I+ LL
 Sbjct: 721 LQH QADVNAKTKLGYSPHLHQAQQGHTDIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11
 Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCL 796
 + T++ G++ LH+AAL G+ ++R L+ +GAN A++ PL++A Q+ H +VVK L
 Sbjct: 71 LETTTFKKGNTALHIAALAGQDEVRELNVYGANVNAQSQRKFTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKNGNTALHEAVIEKHV 856
 L++ A N G TPL A GH +VA L+ +G ALH A
 Sbjct: 131 LENGANQNVATEDGFTPLAVALQQGHENVVVAHLINYGTK----GKVRPALHIAARNDDT 186

Query: 857 FVVELLLH GASVQVLNKRQRTAVDCAE--QNSKIMELL 893
 +LL + + VL+K T + A +N + +LL
 Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29
 Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASA 522
 GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A
 Sbjct: 503 GHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYGVKVRVAELLERDAHP 562

Query: 523 EVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARWGYQG VIE 582
 NG TPLH+A + + D VK L+ S N G TPLHIAA+ V
 Sbjct: 563 NAAGKNGLTPLHVAVHHNNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQMQVEVAR 619

Query: 583 TLLQNGASTEIQNRLKETPLKCA 605
 +LLQ G S ++ TPL A
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28
 Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLL 797
 N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL
 Sbjct: 662 NLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNAPKNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKNGNTALHEAVIEKHV 857
 A N K G +PL A GH ++V LLL++GAS N ++ G T L A ++

Sbjct: 722 QHQADVNAKTKLGYSPLHQAAQGGHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYIS 781
 Query: 858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917
 V ++L + V ++ V + S P V + DV+E + +E ++
 Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827
 Query: 918 KIRKK 922
 K ++
 Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYGATPLHLACQKGYQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCV 545
 G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V
 Sbjct: 35 GVDINTCNQNGNLGLHLASKEGHVKMVELLHKEIILETTTCKGNTALHIAALAGQDEVV 94
 Query: 546 KALVYDVESCRLDIGNEGDTPHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCA 605
 + LV Y ++ ++KG TPL++AA+ + V++ LL+NGA+ + TPL A
 Sbjct: 95 RELVNY---GANVNAQSQKGFTPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVA 151
 Query: 606 L 606
 L
 Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06
 Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVTFFSRDDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKG 507
 L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G
 Sbjct: 457 LQNAKAVNAKADQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREG 514
 Query: 508 YQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCVKALVYYD----- 552
 + L LL +AS G TPLH+A YG + L+ D
 Sbjct: 515 HVETVLALLEKEASQACMTKKGFTPLHVAAKYKVRVAELLERDAHPNAAGKNGLTPLH 574
 Query: 553 --VESCRLDI-----GNE-----KGDTPHIAARWGYQGVETLLQNGASTEIQNRL 597
 V LDI G+ G TPLHIA+ V +LLQ G S ++
 Sbjct: 575 VAVHHNNLDIVKLLPRGSGPHSPAWNNGYTPHIAAKQNVARSLLQYGGSSANAESVQ 634
 Query: 598 KETPLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSM-SA 656
 TPL A M A LS +Q + +S + ++QE +
 Sbjct: 635 VGTPLHLAAQEGHAE-MVALLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIK 690
 Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEMYRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQ 716
 G + T + L A G+++V++LL+ + D+ +A+ + + PL Q
 Sbjct: 691 GVMVDATTR--MGYTPLHVASHYGNIKLVKFLQH-QADV-NAKTKLGYS-----PLHQ 740
 Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSDQGSSPLHVA 751
 + + + +G N S DG++PL +A
 Sbjct: 741 AAQGGHTDI-VTLLKNGASPNEVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07
 Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVVKCL 796
 V D ++ AA G D L++G + N + LHLA ++GH ++V L
 Sbjct: 5 VGFREDAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVKMVEL 64
 Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGINASNNKGNTALHEAVIEKHV 856
 L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+
 Sbjct: 65 LHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENHL 124
 Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885
 VV+ LL +GA+ V + T + A Q
 Sbjct: 125 EVVKFLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26
 Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKGYQSVTLLLLHYK 519
 + G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y
 Sbjct: 42 NQNGNLGLHLASKEGHVKMVELLHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYG 101
 Query: 520 ASAEVDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEGDTPHIAARWGYQG 579
 A+ Q G TPL++A H + VK L+ ++ E G TPL +A + G++
 Sbjct: 102 ANVNAQSQKGFTPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQGGHEN 158
 Query: 580 VIETLLQNGASTEIQ 594
 V+ L+ G +++
 Sbjct: 159 VVAHLINYGTKGKVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21
Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTLTLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYYDVE 554
AT A + G ++ L H + ++ + NG LHLA GH V L++ ++
Sbjct: 13 ATSFLLRAARSG--NLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSVM 614
L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+
Sbjct: 71 ---LETTTKKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEV 127

Query: 615 E 615
+
Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01
Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVAL 828
+ G R AD A A + G+ L + N + +G L A GH ++V
Sbjct: 4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVVE 63

Query: 829 LLQHGASINASNKGNLALHEAVIEKHVVFVVELLLHGASVQVLNKRQRTAVDCAEQNSK 888
LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +
Sbjct: 64 LLHKEIILETTTKKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 123

Query: 889 I 889
+
Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAV 677
+RRQ+ E VQ + + + Q + + Q ++ +K++R V
Sbjct: 1614 DRRQQGQEEVQAEAKNTFTQVVQGNFQNIPEGQVTEEQFTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDLSGNTPLIYA 817
+D++G T L+YA
Sbjct: 1186 EDITGTTKLVA 1197

Pedant information for DKFZphtes3_1817, frame 2

Report for DKFZphtes3_1817.2

[LENGTH] 1050
[MW] 117013.72
[pI] 6.47
[HOMOL] TREMBL:DMANKY_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,
complete cds. 2e-45
[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
3e-12
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YDR264c] 3e-12
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]
3e-08
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
5e-05
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
5e-05
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04
[BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att
[SCOP] dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12
[EC] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12
[PIRKW] phosphotransferase 1e-19
[PIRKW] nucleus 1e-13

[PIRKW] potassium channel 5e-15
 [PIRKW] early protein 2e-13
 [PIRKW] tumor suppressor 1e-09
 [PIRKW] duplication 1e-14
 [PIRKW] tandem repeat 1e-19
 [PIRKW] heterodimer 1e-14
 [PIRKW] potassium transport 5e-15
 [PIRKW] cell cycle control 1e-10
 [PIRKW] serine/threonine-specific protein kinase 1e-19
 [PIRKW] transmembrane protein 5e-15
 [PIRKW] transport protein 5e-15
 [PIRKW] DNA binding 2e-11
 [PIRKW] oncogene 1e-08
 [PIRKW] ATP 1e-19
 [PIRKW] protein kinase inhibitor 1e-09
 [PIRKW] voltage-gated ion channel 5e-15
 [PIRKW] phosphoprotein 4e-38
 [PIRKW] apoptosis 1e-19
 [PIRKW] liver 4e-09
 [PIRKW] integrin binding 3e-16
 [PIRKW] differentiation 2e-12
 [PIRKW] transforming protein 1e-08
 [PIRKW] alternative splicing 1e-40
 [PIRKW] coiled coil 1e-14
 [PIRKW] peripheral membrane protein 2e-38
 [PIRKW] transcription factor 4e-16
 [PIRKW] transcription regulation 2e-16
 [PIRKW] nucleotide binding 5e-15
 [PIRKW] phosphoric monoester hydrolase 1e-12
 [PIRKW] cytoskeleton 8e-39
 [PIRKW] calmodulin binding 1e-19
 [PIRKW] smooth muscle 1e-12
 [SUPFAM] ankyrin 1e-40
 [SUPFAM] death-associated protein kinase 1e-19
 [SUPFAM] ankyrin repeat homology 1e-40
 [SUPFAM] protein kinase homology 1e-19
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07
 [SUPFAM] int-3 transforming protein 1e-08
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38
 [SUPFAM] notch protein 2e-12
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13
 [SUPFAM] rel homology 2e-11
 [SUPFAM] EGF homology 2e-12
 [PROSITE] ATP_GTP_A 1
 [PFAM] Ank repeat
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.05 %

SEQ MALYDEDELLKNPFYLLALQKCRPDLCSKVAQIHGIVLPCKGSLSSSIQSTCQFESYILIP
 SEG
 lawCB

 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVPILFEETFYNEKEESFSILCIAHPLEKR
 SEG
 lawCB

 SEQ ESSEEPLAPSDPFSLKTIEDVREFLGRHSEFRDNIA SFHRTFRE CERKSLRHHIDSANA
 SEG
 lawCB

 SEQ LYTKCLQQLLRDShLKM LAKQEAQMNLKQAVEIYVHHEIYNLIFKYVGTMEASEDAAFN
 SEG
 lawCB

 SEQ KITRSLQDLQQKDIGVKPEFSFNIPRAKRELAQLNKCTSPQQLVCLRKVVQLITQSPSQ
 SEG
 lawCB

 SEQ RVNLETMCADDLLSVLLYLLVKTEIPNWMANLSYIKNFRFSSLADELGYCLTSFEAAIE
 SEG
 lawCB

 SEQ YIRQGSLSAKPPESEGFDRLFLKQRM SLLSQTSSPTDCLFKHIASGNQKEVERLLSQE
 SEG
 lawCB

 SEQ DHDKDTVQKMCHPLCFCDCEKLVSGRLNDPSVVTPFSRDDRGHTPLHVAAVCGQASLID
 SEG
 lawCB


```

SEQ    LLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNGNTPLHLACTYG
SEG    .....
lawCB  .....

SEQ    HEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGYQGVIETLLQNGASTEIQNRLKET
SEG    .....
lawCB  .....

SEQ    PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQ
SEG    .....
lawCB  .....

SEQ    EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKC
SEG    .....
lawCB  .....

SEQ    APAQKRLAKVPASGLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVP
SEG    .....
lawCB  .....

SEQ    LHLACQQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASN
SEG    .....
lawCB  .....

SEQ    NKGNTALHEAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCV
SEG    .....
lawCB  .....

SEQ    ASLDDVAETDRKEYVTVKIRKKWNSKLYDLPEPFTFRQFYFVHSAGQFKGKTSREIMARD
SEG    .....
lawCB  .....

SEQ    RSVPNLTEGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPGLTQTGPGHRRMLR
SEG    .....
lawCB  .....

SEQ    RHTVEDAVVSQGPAAAGPLSTPQEVASRS
SEG    .....
lawCB  .....

```

Prosites for DKFzphes3_1817.2

PS00017 945->953 ATP_GTP_A PDOC00017

Pfam for DKFzphes3_1817.2

```

HMM_NAME      Ank repeat
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLH+AA ++ ++++LL+++GA +N
Query          463  GHTPLHVAAVCGQASLIDLLVSKGAMVN      490

32.12 (bits) f: 496 t: 523 Target: dkfzphes3_1817.2 similarity to ankylins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G TPLH+A++ + ++ LLL + A+
dkfzphes3      496  GATPLHLACQKGYQSVTLLLLHYKASAE      523

Query          f: 529 t: 556 Target: dkfzphes3_1817.2 similarity to ankylins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLH+A+ Y+++++V+ L+ +
Query          529  GNTPLHLACTYGHEDCVKALVYYDVESC      556

42.65 (bits) f: 565 t: 592 Target: dkfzphes3_1817.2 similarity to ankylins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLHIAAR + +++ LLQ+GA+
dkfzphes3      565  GDTPLHIAARWGYQGVIETLLQNGASTE      592

Query          f: 744 t: 771 Target: dkfzphes3_1817.2 similarity to ankylins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G +PLH+AA +++ +++RLL+HGA+
Query          744  GSSPLHVAALHGRADLIRLLKKGANAG      771

```


36.38 (bits) f: 777 t: 804 Target: dkfzptes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

Query *GyTPLHIAARyNNvEMVrLLQHGADIN*
PLH+A++++ ++V+ LL+ +A +N
dkfzptes3 777 QAVPLHLACQQGHFQVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzptes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

HMM *GyTPLHIAARyNNvEMVrLLQHGADIN*
G+TPL++A+ ++ E+V LLLQHGA+IN
Query 810 GNTPLIYACSGGHELVALLLQHGASIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzptes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

Query *GyTPLHIAARyNNvEMVrLLQHGADIN*
G+T+LH A+++ +V +V+LLL HGA++
dkfzptes3 843 GNTALHEAVIEKHVFVVELLLHGASVQ 870

DKF2phtes3_19f19

group: testes derived

DKF2phtes3_19f19 encodes a novel 254 amino acid protein with weak similarity to *S. cerevisiae* protein YFL046w.

The protein contains a RGD cell attachment site.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp

Poly A stretch at pos. 1367, no polyadenylation signal found

```

  1 GGGACACCGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTC CCAGCATTCG TGTGGTGCCA GAACTACCTT GCCCGAAAAGC
101 CTGTGCGGAGA TTTACCCCGT CTTCCGCCTC CCTCCCACCG GAAAACCTCTG
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCGCGGGG CATTTCTCGC CGGCCCTGCG
251 GAGAGAGTTC TTCACTACCA CAACCAAGGA GGGATATGAT AGGCGGCCAG
301 TGGATATAAC TCCTTTAGAA CAAAGGAAAT TAACTTTTGA TACCCATGCA
351 TTGGTTCAGG ACTTGGAAAC TCATGGATTT GACAAAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAACTG CTTTATCAAA TGTGAGCCTG GATACTATCT
451 ATAAAGAGAT GGTCACTCAA GCTCAACAGG AAATAACAGT ACAACAGCTA
501 ATGGCTCATT TGGATGCTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAAG
551 TGAATTGCA AATCTGAGAG CAGAGAATGA GAAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCAACAACCTA ATGCATGAAA CCAGTCGAAT CAGAGCAGAT
651 AATAAACTGG ATATCAACTT AGAAAGGAGC AGAGTAACAG ATATGTTTAC
701 AGATCAAGAA AAGCAACTTA TGGAAACAAC TACAGAATTT ACAAAAAGG
751 ATACTCAAAC CAAAAGTATT ATTTCAGAGA CCAGTAATAA AATTGACGCT
801 GAAATTGCTT CCTTAAAAAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTTTACTTG CCTGGCAATA GCATTGGGAT
901 TTTATAGATT CTGGAAGTAG TATTAATGCT CATCCTGCTG TGGCTGTTGG
951 CTTCTTAGAA CACCAAACCG GGAGAGATTT ACTTTGAACA TTGTCAGTTG
1001 CAGCAAAAAT TTACTACACA AGATTATTCG AAGTGTATAC GGAATAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATACT GTGCTTTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTGTGAGAT AGAAGAGCAT TTTGTCCTTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAT CATAACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAC CATTCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATAATTA CATTTTCTA
1351 TAAATGAAA GATTATTACA ACAAAAAAAA AAAAAAAA AAAAA
```

BLAST Results

Entry HS419346 from database EMBL:
human STS WI-13569.
Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:
human STS SHGC-50338.
Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:
human STS WI-13893.
Score = 1578, P = 1.0e-64, identities = 358/397

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 917 bp; peptide length: 254
 Category: similarity to unknown protein
 Classification: no clue
 Prosite motifs: RGD (15-18)

```

1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTT KEGYDRRPVD
51 ITPLEQRKLT FDTHALVQDL ETHGFDKTQA ETIVSALTAL SNVSLDTIYK
101 EMVTQAQQEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMKIELDQ
151 VKQQLMHETS RIRADNKLDI NLERSRVTDI FTDQEKQLME TTEFTKKDT
201 QTKSISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLAIALGFY
251 RFWK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19f19, frame 3

SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 138, P = 5.4e-08

>SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.
 Length = 211

HSPs:

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09
 Identities = 34/121 (28%), Positives = 67/121 (55%)

```

Query: 70 LETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQE-ITVQQLMAHLDIAIRKDMVI 128
      LE G+ AETI + + ++ +L + K + +A+QE ++ QQ L IRK +
Sbjct: 46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAQESVSFQQKRTFLQ-IRKYLET 104

Query: 129 LEKSEFANLRAENNEKMKIELDQVKQQLMHETSRIADNKLDINLERSRVTDIMFTDQEKQL 188
      +E++EF +R ++K+ E+++ K L + ++ +L++NLE+ R+ D T + +
Sbjct: 105 IEENEFDKVRKSSDKLINEIEKTKSSLREDVKTALSEVRLNLEKGRMKDAATSRNTNI 164

Query: 189 ME 190
      E
Sbjct: 165 HE 166

```

Pedant information for DKFZphtes3_19f19, frame 3

Report for DKFZphtes3_19f19.3

```

[LENGTH] 254
[MW] 29505.73
[pI] 6.99
[HOMOL] PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
[FUNCAT] 99 unclassified proteins (S. cerevisiae, YFL046w) 8e-12
[PROSITE] RGD 1
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 5.12 %
[KW] COILED_COIL 11.02 %

```

```

SEQ MNSRQAWRLF LSQGRGDRWVSRPRGHFSPA LRREFFTTT KEGYDRRPVDITPLEQRKLT
SEG .....
PRD cccchhhhhhhhhccccceeeccccchhhhhhhheeeccccccccccccchhhhhhhcc
COILS .....
MEM .....

SEQ FDTHALVQDLETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQEITVQQLMAHLD
SEG .....
PRD chhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```



```
MEM .....  
  
SEQ AIRKDMVILEKSEFANLRAENEMKIELDQVKQLMHETSRIRADNKLIDINLERSRVTDM  
SEG .....  
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
COILS .. CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC .....  
MEM .....  
  
SEQ FTDQEQLMETTTFTKKDTQTKSIISETSNKIDAEIASLKTLMESNKLETIRYLAASVF  
SEG ..... xxxxxxxxxxxxxxxx .....  
PRD hhhhhhhhhhhhhhhccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
COILS ..... MMMMMMM .....  
MEM .....  
  
SEQ TCLAIALGFYRFWK  
SEG .....  
PRD hhhhhhhhhhhccc  
COILS .....  
MEM MMMMMMMMMM....
```

Prosites for DKFZphtes3_19f19.3

PS00016 15->18 RGD PDOC00016

(No Pfam data available for DKFZphtes3_19f19.3)

DKFZphtes3_19j17

group: testes derived

DKFZphtes3_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
Poly A stretch at pos. 2740, no polyadenylation signal found

```

1 ATTCTCAGCC AAATTTTTT ATTTTTGCA GAATCAGTGT GCAAGGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTTTTGT GTTTAGTGTG ATTTGTTATC
101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTTTGTG TGAGAAACTT
151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAT TTTTAATTTT GTAAGTTCAT
201 AAAATTTAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
251 GGTTTCAATT CTTTCTCTTT TTTTTTGGG GGGTGATGTT TTACAGGCAC
301 TTAAGTATTC ATCGAAGAGT CACCCCAAGT GCGGTGATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAAAATAAAA TGTTGCGGAG
401 ATCTGATAGT CCTGAAAACA AATACAGTGA CAGCACAGGT CACAGTAAGG
451 CCAAAAATGT GCATACTCAC AGAGTTAGAG AGAGGGATGG TGGGACCACT
501 TACTCTCCAC AAGAAAATTC ACACAACCAC AGTGCTCTTC ATAGTTCAAA
551 TTCACATTCT TCTAATCCAA GCAATAACCC AAGCAAAACT TCAGATGCAC
601 CTTATGATTC TGCAGATGAC TGGTCTGAGC ATATTAGCTC TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG AACAGAAGTT TCACAATGGG AAAAACCAAA
701 AGAGTGGCTT GAAAGAGAAC AGAGACAAA AGAAGCAAA AAGATGGCAG
751 TCAACAGCTT CCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGAATG GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTTT GTCTCAAACA AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CCAGTACAGC ACCCCATCAA ACCAGTGGT CATCCAAGT CTACCCCAAG
1001 CACTGTTTCT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTTGA TGCTAATGGA GCATCTACTT TATCAAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAAACAGAA AGAAAAGAA CTACATCAGG
1151 AGACAAACCC GTATCACATT CTGACACAAC TCCTTCCAG TCTTCTGCCT
1201 CTGGACTGAA CCCACATCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTTCTCTCT TCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCCTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTTCTTA CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTT TCAAGCTGCT CAGCTCTCTA
1501 CACAAGATAT CCCTCTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAAT GGGGAAGTGA AGGGTCACTT TGTCAGAAAG CTGATAAACA
1601 GCAGGAATGC CTTGTCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCTCTTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCAGCC ATCTAATCAG
1751 TCTCCGATGT CTTTAACATC TGATGCGTCA TCCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACAAACAG AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTTAGTAC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTGACA GTGAGCCACA CAGCAGCCTG TAACTGCTGA
1951 CAAGCAGCAA GGTGATGAAC CTGTCTCTCC TCGAAGTCTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAAATG CAACAGTTGT ACCACAGAAT TCTTCTGCCC GATCCACGTG

```



```

2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGCGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTTGTAC
2251 TGAATTAAAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATTCAAGCAA
2301 CTTTGCAGAG GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCTT
2351 GAAAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCCC AATCTTAACA
2451 TTTTGGAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGAGTC AATTCAGGGG AAAGATACAA
2551 GATTGATTGG TAAAACCCCTT GAAATGTAGA TTTCTTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTGG TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA TGTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA

```

BLAST Results

Entry AC005876 from database EMBLNEW:
Homo sapiens chromosome 10 clone C1F987SK-1188I5 map 10p11.2-10p12.1,
complete sequence.
Score = 2130, P = 0.0e+00, identities = 426/426
12 exons matching Bp 492-2740

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209
Category: questionable ORF
Classification: no clue

```

1 MSLTSDASSP RSYVSPRIST PQNTVPKPK LISTPPVSSQ PKVSTPVVKQ
51 GPVSQSATQQ PVTADKQQGH EPVSPRSLQR SSQRSPSPGP NHTSNSSNAS
101 NATVVPQNSS ARSTCSLTPA LAAHFSENLI KHVQGWPAH AEQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLVRVCEIQ ATLREQRILF LRQIQEILEK
201 LKNQNSFMV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: WW DOMAIN_1 (90-116)
WW_DOMAIN_1 (90-116)

```

1 MRDAGDPSPP NKMLRRSDSP ENKYSDSTGH SKAKNVHTHR VRERDGGTSY
51 SPQENSHNHS ALHSSNSHSS NPSNNPSKTS DAPYDSADDW SEHSSSGKK
101 YYNCRTEVS QWEKPKEWLE REQROKEANK MAVNSFPKDR DYREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSQTs RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFDAANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPQSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAAVT
351 QASLQSIHK FLTAGPSAFN ITSLISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSLC QKADKQQECL VWNGSIMVQR LLQPSG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 3

TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; *Caenorhabditis elegans* cosmid Y40B1A, N = 1, Score = 144, P = 1.8e-09

>TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; *Caenorhabditis elegans* cosmid Y40B1A
Length = 120

HSPs:

Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09
Identities = 30/67 (44%), Positives = 43/67 (64%)

Query: 90 WSEHISSSGKKYYNCRTEVSQWEKPKEW-LEREQRQKEANKMAVNSFFPK---DRDYRRE 145
W+E +SSSGK YYYN +TE+SQW+KP EW E +++ K VN P+ DR Y
Sbjct: 11 WTEQMSSSSGKMYYNKKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query: 146 VMQATATS 153
+ Q +++S
Sbjct: 70 IGQLSSSS 77

Pedant information for DKFZphtes3_19j17, frame 2

Report for DKFZphtes3_19j17.2

[LENGTH] 209
[MW] 22873.85
[pI] 9.95
[KW] All Alpha
[KW] LOW_COMPLEXITY 13.40 %

SEQ MSLTSDASSPRSYVSPRISTPQTNTVPIKPLISTPPVSSQPKVSTPVVKQGPVSQSATQQ
SEG
PRD ccc

SEQ PVTADKQQGHPEVSPRSLQRSSQSPSPGNHTSNSSNASNATVVPQNSSARSTCSLTPA
SEG
PRD ccc

SEQ LAAHFSENLIKHVQGWADHAQASRLREEAHNMGTIHMSEICTELKNLRLSLRVCEIQ
SEG
PRD hhhhhhchhhhhhhccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh

SEQ ATLREQRILFLRQQIKELEKLKNQNSFMV
SEG
PRD hhh

(No Prosite data available for DKFZphtes3_19j17.2)

(No Pfam data available for DKFZphtes3_19j17.2)

Pedant information for DKFZphtes3_19j17, frame 3

Report for DKFZphtes3_19j17.3

[LENGTH] 436
[MW] 47716.62
[pI] 8.71
[HOMOL] TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; *Caenorhabditis elegans* cosmid Y40B1A 6e-08

[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPR152c] 6e-04
[BLOCKS] BL01159 WW/rsp5/WWP domain proteins
[PROSITE] WW_DOMAIN_1 2
[PFAM] WW/rsp5/WWP domain containing proteins
[KW] All Alpha
[KW] LOW_COMPLEXITY 22.48 %


```

SEQ  MRDAGDPSPPNKMLRRSDSPENKYSdstGHSKAKNVHthrvrERDGGTSYSPQENSHNHS
SEG  .....xxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  ALHSSNSHSSNPSNNPSKTSdAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKWELE
SEG  xxxxxxxxxxxxxxxxxxx.....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhh

SEQ  REQRQKEANKMAVNSFPKDRDYRREVmqATATSGFASGMEDKHSSDASSLLPQNILSQTs
SEG  .....
PRD  hhhhhhhhhhhccccccccchhhhhhhhhcccccccccccccccccccccccccccccc

SEQ  RHNDRDYRLPRAETHSSSTPVQHPIKPVVHPTATPSTVPSSPFTLQSDHQPKKSFDANGA
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  STLskLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPTSTSSASGLNPTSAPPTSASAVPV
SEG  xxxxxxxxxxx.....xxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPVPQSPiPPLLDPNLLRQLLPALQATLQLNNSNVDISKINEVLTAAVTQASLQSIHK
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cccccccccccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhh

SEQ  FLTAGPSAFNITSLISQAQLSTQDIPLHEGIQMERDTHRSKWEVKGSLCQKADKQQECL
SEG  .....
PRD  hhccccceehhhhhhhhhccccccccccccccccccccccccccccchhhhhhhccce

SEQ  VWNGSIMVQRLQLQPSG
SEG  .....
PRD  eccccchhhhhcccccc

```

Prosites for DKFZphtes3_19j17.3

PS01159	90->116	WW_DOMAIN_1	PDOC50020
PS01159	90->116	WW_DOMAIN_1	PDOC50020

Pfam for DKFZphtes3_19j17.3

HMM_NAME	WW/rsp5/WWP domain containing proteins		
HMM	*LPsGWEeHWDpsGRpWYYWNHETkTTQWEpp*		
	+ ++W EH++ SG+ YY+N T+ +QWE+P		
Query	86	SADDWSEHISSSGKK-YYNCRTEVSQWEKP	115

DKFZphtes3_lcl

group: signal transduction

DKFZphtes3_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```
1 GCGAAGTGAA GGGTGGCCCA GGTGGGGCCA GGCTGACTGA ATGTATCTCC
51 TAGCTATGGA CTAATAATA CATGGGGGGA AATAACAAG TATTCATGAG
101 GGTGAAAATG TGACCCAGCA GGAAAATTAC AACTATTTTC AATTGACGTT
151 GAATAGGATG AGTCATGGAA TTAAAGTGAT TTACTGAAGA TTATACTACT
201 GGTAGATAGA AGAGCTAAAG AAAGATGGAT ACTATGATGC TGAATGTGCG
251 GAATCTGTTT GAGCAGCTTG TCGCCCGGGT GGAGATTCTC AGTGAAGGAA
301 ATGAAGTCCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCCGTAAA
351 AAGTGGCAGA GGACTGACCA TGAGCTGGGG AAATAACAAG ATCTTTTGAT
401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTAAAGCTG AAGCATGCAC
451 GTAATCAGGT GGATGTAGAG ATCAAACGGA GACAGAGAGC TGAGGCTGAC
501 TGCGAAAAGC TGGAACGACA GATTCAGCTG ATTCGAGAGA TGCTCATGTG
551 TGACACATCT GGCAGCATT CAACTAAGCGA GGAGCAAAA TCAGCTCTGG
601 CTTTCTCAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAAAGA
651 CTATCAACCA TTGATGAATC TGGTTCCATT TTATCAGATA TCAGCTTTGA
701 CAAGACTGAT GAATCACTGG ATTGGGACTC TTCTTTGGTG AAGACTTTCA
751 AACTGAAGAA GAGAGAAAAG AGGCGCTCTA CTAGCCGACA GTTTGTTGAT
801 GGTCCCCCTG GACCTGTAAA GAAAACCTCG TCCATTGGCT CTGCAGTAGA
851 CCAGGGGAAT GAATCCATAG TTGCAAAAAC TACAGTGACT GTTCCCAATG
901 ATGGCGGGCC CATCGAAGCT GTGTCCACTA TTGAGACTGT GCCATATTGG
951 ACCAGGAGCC GAAGGAAAAC AGGTACTTTA CAACCTTGGG ACAGTGACTC
1001 CACCTGTAAC AGCAGGCAGC TGGAGCCAAG AACTGAGACA GACAGTGTGG
1051 GCACGCCACA GAGTAATGGA GGGATGCGCC TGCATGACTT TGTTCCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTCCA TGTGAAAGC GGATAAAAT
1151 TGGCAAATTA TCTCTGAAGT GTCGAGACTG TCGTGTGGTC TCTCATCCAG
1201 AATGTCGGGA CCGCTGTCCC CTTCCCTGCA TTCTACCCCT GATAGGAACA
1251 CCTGTCAAGA TTGGAGAGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCATTG TGTAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCGCACA
1401 GTAAAAGAGC TGAAAGAGAA ATTCTCTAGA GTGAAAAGT TACCCTCTCT
1451 CAGCAAAGTG GATGATATCC ATGCTATCTG TAGCCTTCTA AAAGACTTTC
1501 TTGAAAACCT CAAAGAACCT CTTCTGACCT TTCGCCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAAGTGC CCCAGGCCAA CAGGGACACA TTAGCTTTCC
1651 TCATGATTCA CTTGCAGAGA GTGGCTCAGA GTCCACATAC TAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGGCCCT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCCTGCTT TCCTTGCCCTC TGGAGTATTG GAGTCAGTTC
1851 ATGATGGTGG AGCAAGAGAA CATTGACCCC CTACATGTCA TTGAAAACCTC
1901 AAATGCCTTT TCAACACCAC AGACACCAGA TATTAAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCTGAA CATCAGCTTC TCAAGACTCC TTCCTCTAGT
2001 TCCCTGTGAC AGAGAGTCCG TTCCACCCCTC ACCAAGAACA CTCCTAGATT
2051 TGGGAGCAAA AGCAAGTCTG CCACTAACCT AGGACGACAA GGCAACTTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCCTGTT ACTTCCCAGC
2151 ATTGACTGAC TATAAGAAA GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTACTCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTGCATGAG GGTTTTATTA AAATATATA TATCTCCCTC TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TTCTTTTGG GGAAGGGGGT TATTGTTCCT TTGGCTTAAA GCCAAATGCT
2451 GCTCATAGAA TGATCTTTCT CTAGTTTCAT TTAGAAGTGA TTTCCGTGAG
2501 ACAATGACAG AAACCCCTACC TATCTGATAA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTTAGGAT
```



```

2601 GCCTCCTTCT AAGAACCAGA AGTTCTCATT CCCCATTTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTTGGATGA TTAGGTCTTT AATAGTGTG
2751 AGTGGCACAA CCTTGTAAT GTGAAAGTAC AACTCGTATT TATCTCTGAT
2801 GTGCCGCTGG CTGAACCTTG GGTTCATTG GGGTCAAAGC CAGTTTTTCT
2851 TTTAAAATTG AATTCATTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCAAGTGGAGC CCAACTTCTA AAGGTCAATA TATCATCCTT TGGCATCCCA
2951 ACTAACAATA AAGAGTAGGC TATAAGGGAA GATTGTCAAT ATTTGTGGT
3001 AAGAAAAGCT ACAGTCATTT TTTCTTTGCA CTTGGATGC TGAATTTTT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTTA
3101 AAATTATTCT TAATGTCTGT AAAAAAGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCTGATTG ACCCTTATCT GTAAAACACC TATTGGGAT AATATTGGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAATGAAA AAAAAA

```

BLAST Results

Entry U82984 from database EMBLST:
Homo sapiens DRES 56 mRNA sequence.
Score = 8775, P = 0.0e+00, identities = 1757/1758
matches 3' end

Medline entries

93074974:
Developmental regulation and neuronal expression of the mRNA of rat
n-chimaerin, a
p21rac GAP:cDNA sequence.

93024458:
A Drosophila rotund transcript expressed during spermatogenesis and
imaginal disc
morphogenesis encodes a protein which is similar to human Rac
GTPase-activating
(racGAP) proteins.

Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632
Category: similarity to known protein

```

1 MDTMMLNVRN LFEQLVRRVE ILSEGNEVQF IQLAKDFEDF RKKWQRTDHE
51 LGKYKDLLMK AETERSALDV KLKHARNQVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIQLSEE QKSALAFLNR GQSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLDWDSS LVKTFKLKKR EKRRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNESIVAKTT VTPNDGGPI EAVSTIETVP YWTRSRRTG
251 TLQPWNDSST LNSRQLEPRT ETDSVGTPOS NGGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKCRCR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFVSQTSP MIPSIVVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRNLN AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPVTMLQD IKRQPKVVER LLSLPLEYWS QFMMVEQENI
551 DPLHVIENSN AFSTPQTPDI KVSLLGPVTT PEHQLLKTPS SSSLSQVRVS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK

```

BLASTP hits

Entry CEK08E3.4 from database TREMBLNEW:
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit
fly (Drosophila melanogaster) (fragment)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit
fly (Drosophila melanogaster)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539_1 from database TREMBL:
 gene: "rotund"; product: "rnracGAP"; *Drosophila melanogaster* rnracGAP
 (rotund) gene, complete cds.
 Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:
 N-chimerin - rat
 Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKFZphtes3_1c1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_1c1, frame 3

Report for DKFZphtes3_1c1.3

[LENGTH] 632
 [MW] 71026.84
 [pI] 9.08
 [HOMOL] PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -
 fruit fly (*Drosophila melanogaster*) 2e-46
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]
 2e-11
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR127w] 5e-09
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08
 [FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08
 [BLOCKS] BL00479B Phorbol esters / diacylglycerol binding domain proteins
 [BLOCKS] BL00479A Phorbol esters / diacylglycerol binding domain proteins
 [SCOP] dlpbwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo sapiens)] 1e-55
 [SCOP] dlrgp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 1e-49
 [PIRKW] breakpoint cluster region 1e-19
 [PIRKW] transmembrane protein 7e-08
 [PIRKW] brain 3e-22
 [PIRKW] alternative splicing 1e-19
 [PIRKW] P-loop 2e-25
 [SUPFAM] CDC24 homology 3e-22
 [SUPFAM] bcr protein 3e-22
 [SUPFAM] myosin motor domain homology 2e-25
 [SUPFAM] pleckstrin repeat homology 4e-10
 [SUPFAM] LIM metal-binding repeat homology 2e-09
 [SUPFAM] protein kinase C zinc-binding repeat homology 5e-29
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 1
 [PROSITE] DAG_PE_BINDING_DOMAIN 1
 [PFAM] Phorbol esters / diacylglycerol binding domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 2.22 %
 [KW] COILED_COIL 8.54 %

SEQ MDTMMLNVRNLFQVLRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK
 SEG
 COILSCCCCCCCCCCCC
 lrgp-
 SEQ AETERSALDVKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSIQLSEE
 SEG
 COILS CC
 lrgp-
 SEQ QKSALAFNLRGQPPSSNAGNKRSLTIDESGSIILSDISFDKTDSELDWSSLVKTFKLKKR
 SEG
 COILS

Prosites for DKFZphtes3_1c1.3

661

PS00479 287->336 DAG_PE_BINDING_DOMAIN PDOC00379

Pfam for DKFZphtes3_1c1.3

HMM_NAME	Phorbol esters / diacylglycerol binding domain		
HMM	*HrFmrHTFrqPTWCDHCgeFIWGWgKQGYQCQnCgMNCHKRChelVPmm		
	H+F+ +T + P +C CG +I +GK ++C +C+++ H +C+ + P		
Query	287	HDFVSKTVIKPESCVPCGKRI-KFGKLSLKCRDCRVVSHPECRDRCPLP	334
HMM		C*	
		C	
Query	335	C	335

DKFZphtes3_lgl3

group: intracellular transport and trafficking

DKFZp DKFZphtes3_lgl3 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cp151 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cp151"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cp151 shows haploid-specific transcription!
testis or haploid-specific transcription

Sequenced by DKFZ

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```

1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
51 CATTCTCTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGGCGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTCTGAG GAGTCAGAGG TGGAAATTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAGAAAAAAT
401 TGCTGGTCCT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGCAGACT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCCTAGAGA AGCAGACTTC
501 CGACCTGGTT CTTCTGCACC ATCACTGCAA ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAAATCACA ACGAGAACAC AGGGGAGAAG
601 CTCATTTTGG CGCAGGAGCA ACTCGCCTTG GCCGGGGACA AGATCGCCTC
651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAG TCTTCCTTGA
701 GCAACATCGA GTTACTAGAA TGCCAAAGTA AGATGTTGCA GGGGGAACTC
751 GGCGGGATCA TGGGTCAGGA GCCTGAGAAC AAGGGTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCAACAGGA TGATCTCATT
901 CAAGAACTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
951 TGAAAGGCTT TGTATAAAAC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCCTCCTAGC TCCTCAGAAG AGTGTGAAGA CATCAAAAAG
1051 ATACTGAAAGC ACTTGCAGGA CCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TCGGGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCTGTCAGT GCCGGCTGCA GGAGCTGCAG CTGGAGTTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCCT CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCCTGCAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCCTGGCA GGCTGTCACC TGGAGGACAC CCAGAGGAAA
1651 CTGCAAGAGG GTCTCTCTCT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACTTCAGA TGCTGCAGAA GGAGTCTCG ATGGCTGAGA
1751 AGGAACAAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAAAATTCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCACAGG GAGCAAGGCT CCATCAAATG CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TGCGGCAGGA ATTTAAAAAG AAAGACAAGA CGTTGAAGA GAATTCCAGA
2101 AAGTTGGAGG AAGAAAATGA GAATCTCCGA GCAGAGCTAC AGTGTGTGTC
2151 TACACAACATG GAATCCTCTC TCAACAAATA CAACACCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGCAAGCCC AGCTGGACAA AGCTCTGCAG AAGGAGAAGC ACTATCTCCA
2301 GACTACCATC ACCAAGAAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CCTGCCAGGA TGACCTGACA CAAGCCCTCG AGAAGCTCAA TCACGTGACC
2401 TCAGAGACAA AGAGCCTGCA GCAAAGCTTG ACACAGACCC AAGAGAAGAA

```



```

2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCTTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTCG CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCCAAAAGGA
2901 CCACCTCCAC AGTGTAAATG TCCACTTGCA GCAGGAAAAC AAGAAAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAAGCCC TAGGCCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGCGGCACC TTGGGCTGGA AGGGGTGCCC CCAGGATATG GGTCAAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCCTCATA
3151 TGCTAGAAAT CACATCTAGC CCTGAGCAGC ATTTCCACGG GTGTTTCTTC
3201 AGAGGACAGT GAGTTCACAG CCCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCACGGT CCCAGCCTAT TTGCAAGAC ACTAACTTTT
3301 GTTGAGTTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACCT TGAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA

```

BLAST Results

Entry AC004682 from database EMBLNEW:
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete
sequence.

Score = 1291, P = 0.0e+00, identities = 265/272

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007

Category: similarity to known protein

Prosite motifs: LEUCINE_ZIPPER (83-105)

LEUCINE_ZIPPER (90-112)

LEUCINE_ZIPPER (97-119)

LEUCINE_ZIPPER (104-126)

LEUCINE_ZIPPER (403-425)

LEUCINE_ZIPPER (410-432)

LEUCINE_ZIPPER (918-940)

```

1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQQLKKKL LVLQEELEFH
101 TEELQTSYYS LRQYQSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
151 ENTGEKLHLA QEQLALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
201 MLQGELEGIM GQEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVWQKVS
251 QQDDLIQELR NKLACSNAIV LEREKALIKL QADFASCTAT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQKRNIMKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLLKKD
401 KFLQEKDEML QELEKKLTQV QNSLLKKEKE LEKQQCMATE LEMTVKEAKQ
451 DKSKEAECKA LQAEVQKLKN SLEEAQQQER LAAQQAQCK EEAALAGCHL
501 EDTQRKLQKG LLLDKQKADT IQELQRELQM LQKESSMAEK EQTSNRKRVE
551 ELSLELSEAL RKLENSDKEK RQLQKTVAEQ DMKMNMDLDR IKHQHREQGS
601 IKCKLEEDLQ EATKLLLEDKR EQLKKSKEHE KLMEGELEAL RQEFKKKDKT
651 LKENSRLKEE ENENLRAELQ CCSTQLESSL NKYNTSQQVI QDLNKEIALQ
701 KESLSLQAO LDKALQKEKH YLQTTITKEA YDALSRSKSA CQDDLTOALE
751 KLNHVTSSETK SLQQSLTQTO EKKAQLEEEI IAYEERMKKL NTELRKLGRF
801 HQESELEVHA FDKKLEEMSC QVLQWQKQHQ NDLKMLAAKE EQLREFQEEM
851 AALKENLLED DKEPCCLPQW SVPKDTCLRY RGNDQIMTNL EQWAKQOKVA
901 NEKLGNLQLE QVNYIAKLSG EKDHLSVMV HLQENKKKL KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGLGWK GLPQDMGQRM DLTKYIGMPH
1001 CPGSSYC

```

BLASTP hits

Entry HS417401_1 from database TREMBL:
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete

cds.

Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862

Entry SCINTANA_1 from database TREMBL:

Saccharomyces cerevisiae integrin analogue gene, complete cds.

Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897

Entry HS6802_2 from database TREMBL:

gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS.

Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028

Entry AF092090_1 from database TREMBL:

product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.

Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

Alert BLASTP hits for DKFZphtes3_lgl3, frame 1

TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin, N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401_1 product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA_1 Saccharomyces cerevisiae integrin analogue gene, complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin Length = 2,185

HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34
Identities = 212/816 (25%), Positives = 420/816 (51%)

```

Query:   145 EMGNHNEN-TGEKLHLAQEQLALAGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQ 203
          +M + E+ G L +EQL ++ +ERSL+ YR KY ++ ++L+ + K LQ
Sbjct:   119 DMDSEAEEDLVGNSDSLNEQLI---QRLRRMERSLSSYRGKYSSELVTAYQMLQREKKKLQ 175

Query:   204 GELGGIMGQEPENKGDHSGVRIYTPSPMIQEHQETQKRLSEVWQ-KVSQDDLIQELRNK 262
          G I+ Q D S RI +Q Q+ +K L E + + ++D I L+ +
Sbjct:   176 G-----ILSQSQ----DKSLRRIAELREELQMDQQAQKHLQEEFDASLEEKDQYISVLQQT 227

Query:   263 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKHLQE 313
          ++ + + ++ K L +L+ A P S E ED K L+ LQ+
Sbjct:   228 VSLKQRLRNGPMNVDVLKPLPQLEPQ-AEVFTKEENPESDGEFVVDGTSVKTLETLLQ 286

Query:   314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQKRNIMKDMMLKLELDLHGLREETS 366
          + Q C ++ ++ L E EA+ EQ ++++ K++ DLH + E+T
Sbjct:   287 RVRKQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERLQLEKIK-DLH-MAEKTKL 344

Query:   367 HIERKDKDITILQCRQLQELQLEFTTETQKLTLLKKDKFLQEKDEMLQLEKKLTVQ--QNSL 424
          + +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q
Sbjct:   345 ITQLRDAKNLIEQLE-QDKGMVIAETKR---QMHTLEMKEEEIAQLRSRIKQMTTQGE 400

Query:   425 LKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKNSLEEAKQOERLAAQ 484
          L+++KE ++ ELE + A+ K++EA K L+AE+ + ++E+ ++ER++ Q
Sbjct:   401 LREQKE-KSERAFAEELEKALSTAQ--KTEEARRK-LKAEMDEQIKTIEKTSEEERISLQ 456

Query:   485 QA-AQCKEEAA-LAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKESSMAEKEQ 542
          Q ++ K+E + E+ KLQK L +K+ A QEL ++LQ ++E E+ +
Sbjct:   457 QELSRVQEVVDVMKKSSEEQIAKLQK--LHEKELARKEQELTKKLQTRERE--FQEQMK 512

Query:   543 TSNRKRVEELSLELSEALRKLENSDKERQLOKT--VAEQDMKMNDMLDRIKHQHQREQGS 600
          + K E L++S+ + E+ E+ +LQK + E + K+ D+ +
Sbjct:   513 VALEKSQSEY-LKISQEQEQESLALLEELELQKKAILTESENKLRDLQQAETRYRTRILE 571

Query:   601 IKCKLEEDLQEATKLEED----KREQLKKSKEHEKLMEG---ELEALR-QEFKKKDKTL 651
          ++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L
Sbjct:   572 LESSLEKSLQENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESKHKQDQDALWTEKL 631

Query:   652 KENSRKLEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706
          + ++ + E E LR + C + E+ L +K Q I+++N++ + +++ L S
Sbjct:   632 QVLKQQYQTEMEKLRK---CEQEKETLLKDKIIFQAHIEEMNEKTLEKLDVVKQTELES 688

Query:   707 LQALQDKALQKEKHLYLT--TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQ 764
          L ++L + L K +H L+ ++ K+ D + ++ A D+ Q V S K +

```


Sbjct: 689 LSSELSEVL-KARHKLEELSVLKDQTDKMKQELEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRLKRGFHEQSELEVHAFDCKLEEMSCQVLQ 824
S+ +T+ KA L+++I E +K+ + L++ + + E ++ + +L++ S ++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLKERDKHLKEHQAHVENLEADIKRSEGELQQASAKLDV 802

Query: 825 WQKQHNDLKMMLAAKEEQLEFQEEMAALKENLLEDDKEPCCLPQW-----SVPKDTIC-R 878
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + KD C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLQKLLDLETERILLTKQVAEVEAQKQDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQKQVANEKLGQLEQVNYIAKLS-GEKDLHLSVMVHLQEQENK 937
L Q+ ++Q KQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKKIQVQDLMQLEKQNSEMEQKVKSLT--QV-YESKLEDGNKEQEQTKQILVEKENM 912

Query: 938 KLK-KEIEEKKMKAENTRLCTK 958
L+ +E ++K+++ +L K

Sbjct: 913 ILQMRGQKKEIEILTQKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLSN-KLL-SLQLDIKNLHDVCKRQKKTLDN-QLCM-----EEAM 51
K+E E D E V S K L +LQ +K ++ KR ++T+Q + + C +EA+

Sbjct: 260 KEENPESDGEVVEDGTSVKLTETLQQRVKRQENLLKRCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSHDKKQAQALAFEESEVEFGSSKQCHLRQ----LQOLK--KKLLVLQOELEFHTTELQ 105
D++ + ++ + + LR ++QL+ K +++ + + + H E L+

Sbjct: 320 QEQLDERLQLEKIKDLHMAEKTKLITQLRDAKNLIEQLEQDKGMVIAETKQRMH-ETLE 378

Query: 106 TSYYSRLRQYSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNTGKELHLAQEQL- 164
+ Q +S +++ T+ L K K + E E +T +K A+ +L

Sbjct: 379 MKEEIEAQLRSRIKQMTTQGEELREQ-KEKSERAFAEELEKAL---STAQKTEEARRLK 434

Query: 165 ALAGDKIASLERSLNLRYDKYQSSLSNI--ELLEQVQKMLQGEELGGIMGQEPENKGDHRSK 222
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 AEMDEQIKTIETSEEERISLQQLSRVKQEVVDVMKKSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYTPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQA 282
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 EQELTKKLQTRE-REFQEOMKVALEK-SQSEYL--KISQKEQQESLALAE-----LELQK 544

Query: 283 DFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAV-SE 341
A T + +E E + + L+ + ++E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKLRDLQQAETRYRTRILELESSLEKS---LQENKNQSKDLAVHLEAEKNK 600

Query: 342 QKRNMKDMKLELDLHGLREETSABIERKDKDITI-LQCRLELQLEFTEQKLTLLKD 400
+ I + K + +L L+ + A K + + Q +++L+ E E +K TL KD

Sbjct: 601 HNKEITVMVEKHKTELES LKHQQDALWTEKLVQKQYQTEMEKLR-EKCEQEKETLLK 659

Query: 401 K-----FLQEKDEM-LQELEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAKQDKS 453
K ++E +E L++L+ K T+++ SL + E+ K + E E+V + + DK

Sbjct: 660 KEIIFQAHEEMNEKTLEKLDVKQTELE-SLSSELSEVLKARHKLEE-ELSVLKDQTDKM 717

Query: 454 K-EAECKALQAEVQKLKNSLEEAKQERLAAQQAQC-KEEAALAGCHLEDTQKRLQKGL 511
K E E K + + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKERDKHLKEHQ 776

Query: 512 L-LDKQKADTIQELQRELMQKLESSMAEKEQTSNRKRVEELSLELSEALRKLNSDKEK 570
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGELQQASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDLETER 835

Query: 571 RQLQKTVAEQDMKMDM---LD--RIKHQHQREQSGIK--CKLEEDLQEATKLEDDKREQL 623
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVEAQKQDVCTELDAHKKIQVQDLMQLEKQNSEMEQKVKSLTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMEGELEALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLN 681
K +E K L+E E L+ +K K ++ ++KL + +++ + T+ ++

Sbjct: 896 NKEQEQTQKILVEKENMILQMRGQK-KEIEILTQKLSAKEDSIHILNEEYETKFKNQEK 954

Query: 682 KYNTSQVQIQDLNKEIALQKESLMSLQAQLDKALQKEKHLYLQTTITKEAYDALSRKSAAC 741
K +Q +++ + + K+ L+ +A+L K L E L+ + ++ +A + A

Sbjct: 955 KMEKVQKAKEMQETL---KKKLLDQEAKLKKEL--ENTALELSQKEKQFNKMLEMAQA 1009

Query: 742 QD-DLTQALEKLNHVTSETKSLQSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLKRGF 800
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQKEQIE-SLTVHRR--ELNDVISIWE---KKLNQQAELQEI 1061

Query: 801 HQESELEVHAFDCKLEEMSCQVLQW--QKQHNDLKMMLAAKEEQLEFQEEMAALKENLL 858
H E+++ +++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLQEKEQVAELKQKILLFGCEKEEMNK-EITWLKEEGVRQ-DTTLNELQEQLK 1116

Query: 859 EDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQ--WAKQQKVANEKLGNLREQVNYI- 915
 + L Q K L + + +L++ + ++Q V + L + + +V+ +
 Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVSILT 1175

Query: 916 AKLSGEKDLHLSVMVHLQEQENKKL-KEIEEKKMAE 951
 +KL + S+ ++ NK L+ K +E KK+ E
 Sbjct: 1176 SKLKTDEEFQSLKSSHEKSNKSLKEDKSLEFKKLSEE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26
 Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQCMEEAMNSSHDKQAQALAFESE 69
 +E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +
 Sbjct: 560 QEATYRTRILELESSLEKSLQENKNSKDLAVHL----EAEKNKHNEKIT--VMVEKHK 613

Query: 70 VEFSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQSILEKQTSIDLVLH 129
 E S K H +Q +KL VL+Q+ + E+L+ Q + L K +++
 Sbjct: 614 TELESK--H-QQDALWTEKLQVLKQYQTEMEKLEK--CEQKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHNTENGEL---HLAQEQLALAGDKIASLERSLNLYRD 183
 H ++ E +++ + + E+ + + E L H +E+L++ D+ +++ L D
 Sbjct: 667 HIEEMNEKTLEKLDVKQTELESLSSELSEVLKARHKEELSVLKDQTDKMKQELEAKMD 726

Query: 184 K----YQSSLSNIELLECQVKMLQGE--LGGIMQEPENKGDHKSRIYTSPCMIQEHQE 237
 + +Q + +I + E +V + + E L + Q + K + ++ +
 Sbjct: 727 EQKNHQQQVDSI-IKEHEVSIQRTEKALKDQINQLELLKERDK-HLKEHQAHVENLEA 784

Query: 238 TQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297
 KR Q+ S + D+ Q ++ ++ E+ L +LQ T R
 Sbjct: 785 DIKRESEGLQOASAKLDVFSQYSQ---ATHEQTKAYEEQLAQQLLDLE-TERIL--- 837

Query: 298 SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVEAVSEQKRNIMKDMML-ELD 356
 + K + ++ QK C ++ ++ V+DL +LE + + +K + ++ E
 Sbjct: 838 -----LTKQVAEVEAQKKDVCTELDAHKIQVQDLMQLEKQNSEMEQKVSLTQVYESK 891

Query: 357 LH-GLREETSASIERKDKDITILQCL-QLQLEFTEETQKLTLLKKDF--LQEKDEM-LQ 411
 L G +E+ +K+ ILQ R Q+ ++E TQKL+ K+D L E+ E +
 Sbjct: 892 LEDGNKEQEQTQKILVEKENMILQMRGQKKEIIL-TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKLTQVQNSLLK-----KEKELEKQCMATELEMTVKEAKQDKSKEAECKALQAEVQ 466
 EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q
 Sbjct: 951 NQEKMKKVKQKAKEMQETLKKLLDQEARLKKELENTALQELSQ-KEKQFNAKMLEM-AQ 1008

Query: 467 KLKNSLEAAKQERLAAQQAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKADTIQELQR 526
 + +A RL Q Q + + L D +K L Q+A+ +QE+
 Sbjct: 1009 ANSAGISDAVS--RLETNQKEQIESLTVHRRRLNDVISIWEKKL---NQAEELQEIH- 1062

Query: 527 ELQMLQKESMAEKEQT-----SNRKR--EELSLESEALRKLNSDKERQLQ 574
 E+Q+ +KE +AE +Q K + +E ++ L +L+ K+K
 Sbjct: 1063 EIQLQEKEQEVAELKQKILLFGCEKEEMNEKITWLKEGVKQDTTLNELQELQKQSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHQHQREQGSIKCKLEEDLQEQATKLEDDKREQLKKSKEHEKLM 634
 ++A+ + K+ L++++ + L+E L E L E+ + ++ + K +
 Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVSILTSLKTTD 1182

Query: 635 GELEALRQEFKKKDKTKLNSRKLEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694
 E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +
 Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLEFKKLSEELAIQLDICKKTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLAQDLKALQKEKHYLQTTITKEAYDALSRKSAACQDDLT----QALE 750
 K A+ + Q+ K KE ++T E +A R+ Q+ L QA
 Sbjct: 1242 KTNAILSR-ISHCQHRRTKV--KEALLICTVSEL-EAQLRQLTEEQNTLNISFQQATH 1297

Query: 751 KLNHVTSKSLQSLTQTEKKAQLEEEIAYEERMKKLN---TELK--LRGFHQESE 805
 +L ++ KS++ + +K L++E ++ + T+L+K + +
 Sbjct: 1298 QLEKENQIKSMKADIESLVEKEALQKEGGNQQAASEKESCITQLKKELSENINAVTL 1357

Query: 806 LEVHAFDCKLE--EMSCQVLQWQKHQNDLKMALAAKEEQLEFQEEMAALKENLLEDDKE 863
 ++ +KK+E +S Q+ Q QN + L+ KE + +++ K LL D +
 Sbjct: 1358 MKEELKEKKVEISSLSKQLTDLNVQLQNSIS-LSEKEAAISSLRKQYDEECELL-DQVQ 1415

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQQKVANEKLGNLRE---QVNYIAKLSG 920
 ++ K+ D +W K+ + + N ++E Q+ +K +
 Sbjct: 1416 DLSFKVDTLSKEKISALEQVDDWSNKFSEWKKQAQSRFTQHQNTVKELQIQLELKSKEY 1475

Query: 921 EKDH-LHSVHVLQEQENKK---LKKEIEEKKMAE 951
 EKD ++ + L Q+NK+ LK E+E+ K K E
 Sbjct: 1476 EKDEQINLLKEELDQONKRFDCIKGEMEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25
 Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQKTLQDNQLCMEAMNS----SHD 56
 MK + E+ ++ L+ K L+ + + + + R+R+ + ++ +E++ + S +
 Sbjct: 470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLQTREREFQEQMKVALEKSQSEYLRKISQE 528

Query: 57 KKQAQALAFESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTEELQTSYYSRLRQYQS 116
 K+Q ++LA EE E++ K+ L + + KL LQOE E + + SL +
 Sbjct: 529 KEQESLALAELELQ---KKAILTESEN---KLRDLQOEAEYRTRILELESSLEKSLQ 581

Query: 117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE-----EMGNHNENT--GEKLHLAQEQLALA 167
 + Q+ DL + K K ++ ++ E+ E H ++ EKL + ++Q
 Sbjct: 582 ENKNQSKDLAVHLEAEKNKHNEITVMVEKHKTELESKHKQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNYRDK---YQSSLS--NIELLECQVKMLQGEELGGIMGQEPENKGDH 220
 +K+ + L +DK +Q+ + N + LE ++ + Q EL + + E
 Sbjct: 642 MEKLRKCEQEKETLLKDKKEIFQAHIEEMNEKTE--KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKL 280
 K+ S +++ +T K E+ K+ +Q + Q+ + + + +R+ + +K
 Sbjct: 701 HKLEEELS--VLKD--QTDKMKQELEAKMDEQKNHHQQVDSIIKEHEVSIQRTKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEA 338
 Q + R + E+++ +K + + ++ +Q+ + +A
 Sbjct: 757 QINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELOQASAKLDVQSYQSATHEQTKA 816

Query: 339 VSEQKRNIMKMMKLELDLHGLREETSABIERKDKDITILQCRQLQELQLEFTTQKLTLLK 398
 EQ + + ++ LE + L ++ A +E + KD+ C EL + Q L +
 Sbjct: 817 YEEQLAQQLQKLLDLETERILLTKQV-AEVEAQKKDV----CT--ELDAHKIQVQDLMOQ 869

Query: 399 KDKFLQEKDEMLQLEKKLTQVQNSLLKK-EKELEKQOCMATELEMTVKEAKQDKSKEAE 457
 +K + EM Q++ K LTQV S L+ KE E+ + + E E + + ++ + KE E
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGNKEQEQTQKILVEKENMILQMREGQKKEIE 925

Query: 458 C--KALQAEVQKLNKNSLEEAQKQERLAAQQAQCKEEAALAGCHLEDTRK--LQKGLLL 513
 + L A+ + EE + + + ++ + K+A +++T +K L + L
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFKNQEKMKVKQKAK---EMQETLKKKLLDQEAKL 981

Query: 514 DKQKADTIQEL-QRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKEKRO 572
 K+ +T EL Q+E Q K MA+ V L E + L ++ +R+
 Sbjct: 982 KKELENTALELSQKEKQFNAKMLEMAQANSAGISDAVSRLNETNQKEQIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDKMKNMDLRIKHQHQREQGSICKLEEDLQEATKLLDKREQLKKS----KE 628
 L ++ + K+N + ++ H Q K + +L++ L ++E++ K KE
 Sbjct: 1040 LNDVISIWEKLNQQAELQEIEHEIQLEKEQEVAELKQKILLFGCEKEEMNKEITWLKE 1099

Query: 629 HEKLMEGELEALRQEFKKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQ 688
 + L L+++ K+K + NS L ++ L+A L+ L SL + Q+
 Sbjct: 1100 EGVKQDTTLNELQEQKQSAHV--NS--LAQDETKLKAHLEKLEVDLNKSLKENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLQAL---DKALQ--KEKHYLQTTITKEA---YDALSRSAA 740
 + +L K + L ++L D+ Q K H ++ + LS + A
 Sbjct: 1156 QLVELKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNKSLEDKSLEFKKLSEE-LA 1214

Query: 741 CQDDL-----TQAL-----EKLNHVTSETKSLOQSLTQTEKKAQLEEEIIAYEERMKKL 790
 Q D+ T+AL E +N +S+T ++ ++ Q + ++E ++ + +L
 Sbjct: 1215 IQLDICCKKTEALLEAKTNELINISSKTNAILSRISHCQHRRTTKVKEALIKTCTVSEL 1274

Query: 791 NTELRLRGFHFQSELEVFHAFDKKLEEMSCQVLQWQKHQNDLKMMLAAKEEQLEFQEM 850
 +LR+L + +LEE Q+ K + D++ L ++E L Q+E
 Sbjct: 1275 EAQLRQLTEEQNTLNISFQOATHQLEEKENQI---KSMKADIESLVTEKEAL---QKEG 1327

Query: 851 AALKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLRE 910
 + +KE C + Q + K+ N +T +++ K++KV L QL +
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE----NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKKLKEIEEKKMAE 951
 Q+ LS ++ + S+ +E +L +++ K +
 Sbjct: 1379 LNVQLQNSISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFHTTEELQTSYYSRLRQYQSIL 118
 Q L E+ +++ S+ LR++ +L+++L + QQ + EE S QY S+L
 Sbjct: 165 QMLQREKKKLQGISLQSQDKSLRRIAELREELQMDQOAKHLQEEDFASLEEKDQYISVL 224

Query: 119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT---GEKL---HLAQEQLALA 167
 + Q S L + + D + + + E+ EN GE + + + L
 Sbjct: 225 QTQVSLKQRLRNGPMNVDVLKPLQLEPQAEVFTKEENPESDGEPPVEDGTSVKTLET 284

Query: 168 GDKIASLERSLNYRDKYQSSLSNIELLECQVKMLQGEELGGIMGQEPENKGDHDKVRIYT 227

++ E L ++ QS LL ++ LQ +L + QE E D + +
 Sbjct: 285 QQRVQRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERL-LELEKIKD---LHMAE 340
 Query: 228 SPCMIQEHQETQKRLSEVWQKVSOQDDLIQELRNKLACSNAVLEREKALIKLQADFASC 287
 +I + ++ + + ++ Q +I E+ ++ L ++ E+ + +L++
 Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKQRM--HETLEMKEEE-IAQLRSRIKQM 394
 Query: 288 TATH---RYPSSSEEC--EDIKKILKHLQEOKDSQCLHVEEYQNLVKDL-----RVE 335
 T R SE E++K L Q+ +++++ E +K + R+
 Sbjct: 395 TTQGEELREQKEKSERAAFEELKALSTAQKTEARRKLKAEMDEQIKTIEKTSEEERIS 454
 Query: 336 LEA-VSEQKRNIMKDMMKL--ELDLHGLREETSABIERKDKDITILQCRLOELQLEFTET 392
 L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E
 Sbjct: 455 LQQLSRVQKQEVV-DVMKKSSEEQIAKLQKLHEKELARKEQLTK---KLQTREREFQEQ 510
 Query: 393 QKLTLLKKDKFLQEKDEMLQELEKLTQVQNSLLKKEKELEKQCMATELEMTVKEAKQDK 452
 K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+
 Sbjct: 511 MKVALEKQ--SEYKISQEQE---QESLALLEELELQKKAIL-TESENKLRDLQEQ- 561
 Query: 453 SKEAECKALQAEVQKLKNSLEEAKQER-----LAAQQAQCKEEAALAGCHLEDTQR-K 506
 ++ + L+ E L+ SL+E K Q + L A++ KE + H + + K
 Sbjct: 562 AETYRTRILELE-SSLEKSLQENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESK 620
 Query: 507 LQKGLLLDKQKADTIQELQRELQMLQKESMAEKQTSNRKRVEELSLELSEALRK-LEN 565
 Q+ L+ ++ Q+ Q E++ L +E EKE K + + E K LE
 Sbjct: 621 HQQDALWTEKLQVLQKQYQTEMEKL-REKCEQEKETLLKDKIIFQAHIEMNEKTLEK 678
 Query: 566 SDKEKRQLQKTVAEQDMKMDMLDRIKHQHREQSGI-KCKLEEDLQEA-TKLEDDR--E 621
 D ++ +L+ +E +++++L + +H+ E+ S+ K + ++ QE K+ E K +
 Sbjct: 679 LDVKQTELESLSSE----LSEVL-KARHKLEELSVLKDQTDKMKQELEAKMDEQKNHHQ 733
 Query: 622 QLKKS--KEHEKLMGELEALRQEFKKDKTKENSRLKEEN---ENLRAELQCCSTQL 676
 Q S KEHE ++ +AL+ + + + LKE + L+E ENL A+++ +L
 Sbjct: 734 QQVDSIIKEHVSIRTEKALKDQINQLELLKDKHLKEHQAHVENLEADIKRSEGEL 793
 Query: 677 ESSLNKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSR 736
 + + K + Q +++ +E L LQ +L L+ E+ L TK+ + ++
 Sbjct: 794 QQASAKLDVFSYQSATHEQTKAYEEQLAQLQKQL-LDLETERILL----TKQVAEVEAQ 848
 Query: 737 KSAACQD-----DLTQALEKLNHVTSETKSLQSSLTQTQEKKAQ--LEEEIIAYEE 785
 K C + DL Q LEK N SE + +SLTQ E K + +E+ +
 Sbjct: 849 KKDVTCLDAHKIQVQDLMOQLEKQN---SEMEQVKVSLTQVYESKLEDGKNKEQEQTKQI 905
 Query: 786 RMKKLNTLRLKRGFQHESELEVHAFDKKLEEMSCQVL--QWQKQHQNLDKMLAAKEEQ 843
 ++K N L+ G Q+ E+E+ +E S +L +++ +N K + +++
 Sbjct: 906 LVEKENMILQMRG--QKKEIEILTQKLSAKEDSIHILNEEYETKFNQEKMEKVKQKA 963
 Query: 844 REFQEMAALKENLLEDDKEPCCLPQWVSPKOTCRLYRGNDQIMTNLEQWAKQKQV---- 899
 +E QE LK+ LL+ ++ L + + L + Q + + A+
 Sbjct: 964 KEMQE---TLKKLLDQEAQ---LKK-ELENTALELSQKEKQFNKMLEMAQANSAGISD 1016
 Query: 900 ANEKLGNQLREQVNYIAKLSG-EKDHLHSMVH-LQENKKLKK--EIEEKKMKAENTRL 955
 A +L +EQ+ + ++ E + + S+ L Q+ ++L++ EI+ ++ + E L
 Sbjct: 1017 AVSRLETNQKEQIESLTVHRRELNDVISIWEKKLNQQAELQEIHEIQLQEKEQEVAEL 1076
 Query: 956 CTKALGPSRTESTQREKVCGLTGWKGLPD 985
 K L E + K L +G+ QD
 Sbjct: 1077 KQKIL-LFGCEKEEMNKETIWLKEEGVKQD 1105
 Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25
 Identities = 220/907 (24%), Positives = 444/907 (48%)
 Query: 67 ESEVEFGSSKQCHLRQLQQLKKLLVLQLEFHTTEELQTSYYSRLQYQSILE---KQTS 123
 E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+
 Sbjct: 123 EAEDLVGNSDSLNEQLIQLRRMERSLSSYRGKYSELVTAYQMLQREKKKLQGISLSQSQ 182
 Query: 124 DLVLLHHHCKLKEDEVILYEEEMGNHNTGKHLHLAQEQLALAGDKIASLERSLNLYRD 183
 D L +L+E+ + +++ H + E+ + E+ I+ L+ ++L +
 Sbjct: 183 DKSL-RRIAELREE--LQMDQQAQKHLQ---EEFDASLEE---KDQYISVLQTQVSLKQ 233
 Query: 184 KYQSSLSNIELLECQVKMLQELGGIMGQE-PENKG-----DHSKVR-IYTSPCMIQEHQ 236
 + ++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +
 Sbjct: 234 RLNRGPMNVDLK-PLPQLEPQAEVFTKEENPESDGEPPVEDGTSVKLTLETQQRVQRQE 292
 Query: 237 ETQKRLSEVWQKVSOQDDLIQELRNKLACSNAVLEREKALIKLQADFASCATHRYPSS 296
 KR E Q +Q L+ KA L ER + L K+ D T
 Sbjct: 293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTCLIT-- 346
 Query: 297 SSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELD 356
 + D K +++ L++ K + E + + + L ++ E ++ Q R+ +K M +
 Sbjct: 347 ---QLRDAKNLIEQLEQDKGM---VIAETKQRMHETLEMKEEEIA-QLRSRIKQMTQGE 400

670

AE K Q+ + L+ LEE ++ L Q + + + A +LE+ +QK L
 Sbjct: 1783 AEAQHEDQSMIGHLQEELEEKNKKYSLIVAQHVKEGGKNNIQAKQNLNVFDDVQKTL 1842
 Query: 512 LLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569
 ++K T Q L++++ L +S + +++ +R +EEL+ E +AL++++ +K
 Sbjct: 1843 ---QEKELTQCILEQKIKEL--DSCLVRQKEV-HRVEMEELTSKYELQALQMDGRNKP 1896
 Query: 570 KRQLQKTVAEQD---MKMNDMLDRIKHQHQREQSGIKCKLEEDLQEATKILLEDKREQLKK- 625
 L++ E+ + +L ++ QH + E + Q+ K + ++ L+
 Sbjct: 1897 TELLEENTEEKSKSHLVQPKLLSNMEAQHNDFEKLAGAEREKQKLGKEIVRLQKDLRML 1956
 Query: 626 SKEHEKLMGELEALRQEFKKDKTKLKNRSRKLEENENLRAELQCCSTQLESSLNKYNT 685
 KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT
 Sbjct: 1957 RKEHQ---ELEILKKEYDQ-----EREKIKQEEDL--ELKHNT-LKQLMREFNT 2003
 Query: 686 S-QQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDD 744
 Q Q+L I ++A+L ++ Q+E + L I E D L R +A ++
 Sbjct: 2004 QLAQKEQELEMTIKETINKAQVEAEELSHQEETNQLLKKIA-EKDDDLKR-TAKRYEE 2061
 Query: 745 LTQALEKLNHVTSKSLQSLTQTQEKKAQ-LEEEIAYEERMK--KLNTLRKLRGFH 801
 + A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +
 Sbjct: 2062 ILDAREE--EMTAKVRDLQTQLEELQKKYQKLEQEENPGNDNVTIMELQTQLAQKTTLI 2119
 Query: 802 QESELEVHAFDKKLEEMSCQVLQWQK 827
 +S+L+ F +++ + ++ ++K
 Sbjct: 2120 SDSKLKEQEFREQIHNLEDRLKKYEK 2145

 Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24
 Identities = 213/977 (21%), Positives = 454/977 (46%)
 Query: 4 EAGERD-REVSSLNSKLLSLQLD-IKNLHCVCKRQRKTLQDNQLCMEEAMNSSHDKKQQAQ 61
 E R+ +V S+ K L+ Q + ++ +H++ + Q K + +L + + ++ +
 Sbjct: 1034 EVHRRELNDVISIWEKKLNQQAELQEIEIHEI-QLQEKEQEVAEKQKILLFGCEKEEMNK 1092
 Query: 62 ALAFESEVEFGSSKQCHLRQLQ-QLKKLL---VLQQE--LEFHTTELQTSYYSRQY 114
 + + + E G + L +LQ QLK+K + Q E L+ H E+L+ +
 Sbjct: 1093 EITWLKEE---GVKQDQTTLNELQEQLKQSAHVNSLAQDETCLKAHLEKLEVDLNSLSKE 1149
 Query: 115 QSILEKQTSIDLVLHHHCKLKEDEV---ILYEEEMGNHNTGEKHLAQEQLALAGDKI 171
 + L++Q +L +L K K E+ + +E +++ EK + + E +L K+
 Sbjct: 1150 NTFLEQELVELKMLAEEDKRVSELTSLKKTDEEFQSLKSSHEKSNKSLEDKSLFKKL 1209
 Query: 172 AS-LERSINLYRDKYSSLS--NIELLECQVKMLQGEGLGGIMQEPENKGDHKSRIYTS 228
 + L L++ K ++ L EL+ L I +++ K +
 Sbjct: 1210 SEELAIQLDICKCKTEALLEAKTNELINISSKTNAILSRI--SHCQHRRTTKVKEALLIK 1267
 Query: 229 PCMIQEHQ-----ETQKRLSEVWQKVSQQ-DDLIQELRNKLACSNAVLREKALIKL 280
 C + E + E Q L+ +Q+ + Q ++ + + + A +LV E+E L
 Sbjct: 1268 TCTVSELEAQLRQLTEEQNTLNISFQQATHQLEEKENQIKSMKADIESLVTEKEA----L 1323
 Query: 281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVS 340
 Q + + + S E C I ++ K L E ++ L EE +K+ +VE+ ++S
 Sbjct: 1324 QKEGGN----QQAAASEKESC--ITQLKKELSENINAVTLMKEE----LKEKKVEISSLS 1373
 Query: 341 EQKRNIMKDMKLELDLHGLREETSAAHERKDKDITILQCRLOEL--QLEFTETQKLT-L 397
 +Q ++ + +L S+ ++ D++ L ++Q+L +++ +K++ L
 Sbjct: 1374 KQLTDLNVQLQN-SISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKVDTLSKEKISAL 1432
 Query: 398 KK-DKFLQEKDEMLOELEKKLTQVQNSLLKKEKELEKQQCMATELEMTV---KEAQDKS 453
 ++ D + + E ++ + + TQ QN++ + +LE + A E + + KE ++
 Sbjct: 1433 EQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELKSKEAYEKDEQINLLKEELDQQN 1492
 Query: 454 KEAECKALQAEVQKLKNSLEEAQQRERLAAQQAQKKEEAALAGCHLE-DTQRKLQKGLL 512
 K +C + E K K +E+ + L +Q A + E + +E ++ ++ K
 Sbjct: 1493 KRFDCLKGEMEDDKSKMEKKESNLETLSQTARIMELEDHITQKTIEIESLNEVLKNY- 1551
 Query: 513 LDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKQ 572
 ++QK +EL ++LQ Q+ + +++ L ++ +LE KE
 Sbjct: 1552 -NQKQDIEHKELVQKLQHFQELGEEKDNRVKEAEEKILTLENQVYSMAELETKKKELEH 1610
 Query: 573 LQKTVAEQDMKMDMLDRIKHQHQREQ-GSIKCKLEEDLQEATKLL----EDKREQLKSKS 627
 + +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK
 Sbjct: 1611 VNLVSKSKEEELKALEDRLSESAAKLAELKRKAQKIAAIKKQLLSQMEKEEQYKKG 1670
 Query: 628 EHEKLMGELEALRQEFKKDKTKLKNRSRKLEE-ENENL----RAELQCCSTQLESSLNK 682
 E EL QE +++ L+E + +E ++E L A+ T+ E + ++
 Sbjct: 1671 ESHL---SELNTKLQEREREVHILEEKLKSVESSESQSETLIVPRSAKNVAAYTEQEEADSQ 1727
 Query: 683 ---YNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSA 739
 T ++ I L + + +KE L+ Q +K H+ +E L A
 Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFMRCQYQERLIKLEHAEA 1785

Query: 740 ACQDDLTQALEKLNHVTSET--KSLQQSLTQTQEKKAQLEEEIAYEERMKKLNTLRL 797
 +D Q++ +H+ E K+ + SL Q + + + I ++ ++ + ++K
 Sbjct: 1786 KQHED--QSM--IGHLQEELEKNKKYSLIVAHVEEGGKNNIQAQNLNVFDDVQKT 1841

Query: 798 RGFHQESELEVHAFDCKLEEM-SCQVLQWQKHQNDLKMALAAKEEQLEFQEEAALKEN 856
 QE EL ++K++E+ SC V Q ++ H+ +++ L +K E+L+ Q+ K
 Sbjct: 1842 L---QEKELTCQILEQKIKELDSCLVRQ-KEVHRVEMEELTSKYEKLOALQQMDGRNKPT 1897

Query: 857 -LLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYI 915
 LLE++ E PK + ++ + L A+++K +KLG ++ +
 Sbjct: 1898 ELLEENTEEKSKSHLVQPKLLSNMEAQHNDEFLKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query: 916 AKLSGE-KDHLHSVVMHLQOENK-KLKKEIEEKKMKAENTRLCTKALGPSRTESTQREK 972
 L E + L + QE + K+K+E E+ ++K +T + + T+ Q+E+
 Sbjct: 1954 RMLRKEHQEILEIKKEYDQEREKIKQEEDLELKHNS--LKQLMREFNTQLAQKEQ 2010

Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22
 Identities = 221/952 (23%), Positives = 441/952 (46%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDKNLHDVCKRQRKTLQDNQL---CMEEAMNSSHD- 56
 +K A E R+VS L SKL + ++L ++ K+L+D L + E + D
 Sbjct: 1160 LKMLAEEEDKRVSELTSLKLTDEEFQSLSSHEKSNKSLEDKSLFKLSEELAIQLDI 1219

Query: 57 --KKQAQALAFESE-VEFGSSK-QCHLRQLQLKKLLVLQLELEFHT---EELQTSYY 109
 KK L + +E + SSK L ++ + + +++ L T EL+
 Sbjct: 1220 CCKTEALLEAKTNELINISSSKTNAILSRISHCQHRTTKVKEALLIKTCTVSELEAQLR 1279

Query: 110 SLRQYQSILEKQTSDLVLLHHCKLKEDEVILYEEEMGNHNENTGEKHLAQE---QLAL 166
 L + Q+ L H + KE+++ + ++ EK L +E Q
 Sbjct: 1280 QLTTEEQNTLNISFQQAT---HQLEEKENQIKSMKADI---ESLVTEKEALQKEGNGQQA 1333

Query: 167 AGDKIASLERSLNLRYDKYQSSLSNIELLECQVKMLQGGIMGQEPENKGDHRSKVIY 226
 A +K E + + + +++ + L++ ++K + E+ + Q + V++
 Sbjct: 1334 ASEK----ESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTD-----LNVQLQ 1384

Query: 227 TSPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKLQADFAS 286
 S + ++ + ++ + D +Q+L K+ + L E+ AL ++ D+++
 Sbjct: 1385 NSISLSEKAAISSLRKQYDEEKCELDDQVQDLSFKV---DTLSKEKISALEQVD-DWSN 1440

Query: 287 CTATHRYPPSS--SEECEDIKKILKHLQEQRDSQCLHVEEYQNLVKD-----LRVE-LE 337
 + + S ++ +K++ L E K + +E NL+K+ R + L+
 Sbjct: 1441 KFSEWKKKAQSRTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQONKRFDCLE 1499

Query: 338 AVSEQKRNM-KDMMKLELDLHGLRE---ETSAHIERKDKDITILQCRLELQLE-LEFTET 392
 E ++ M K LE +L E HI +K +I L L+ Q + E
 Sbjct: 1500 GEMEDDKSKMEKESNLETELKSQTARIMELEDHITQKTIEIESLNEVLKNYNQKQDIEH 1559

Query: 393 QKLTLLKDKFLQ---EKDEMLQELEKLLTQVNSLLKKEKELEKQCCMATELEMTVKEAK 449
 ++L K F + EKD ++E E+K+ ++N + + ELE ++ + ++VK
 Sbjct: 1560 KELVQKLOHQFELGEEKDNRVKEAEKILTLENQVYSMAELETKKKELEHVNLSVK--- 1616

Query: 450 QDKSKEAECKALQAEVQKLNKSLLEEAKQERLAAQQAQCKEEAALAGCHLEDTORKLQK 509
 SKE E KAL+ ++ S + + +R A Q+ A K++ +E+ + + +K
 Sbjct: 1617 ---SKEELKALEDRLES--ESAALAELEKRAEKIAAIAKKQLL---SQMEEKEEQYK 1668

Query: 510 GLLLDKQKADT-IQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRLKLENSDK 568
 G + +T +QE +RE+ +L+++ E Q+ + S + A + E +D
 Sbjct: 1669 GTESHLSELNTKLQEREREVHILEEKLSVSSQSETL--IVPRSAKNVAAYTEQEEADS 1726

Query: 569 E---KRQLQK-TVAEQDMKMND-MLDRIKHQHQREQSGIKCKLEEDLQEATKLEEDKREQ 622
 + K +K +V +++ + +L R+ Q +E+ ++ E Q +L+ K E
 Sbjct: 1727 QGCVQKTYEEKISVLQRNLTEKEKLLQRVG-QEKEE-TVSSHFEMRCQYQERLI--KLEH 1782

Query: 623 LKKSKEHE-KLMEGEL-EALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSL 680
 + +K+HE + M G L E L ++ KK + ++ K E N++A+ LE
 Sbjct: 1783 AE-AKHEDQSMIGHLQEELEKNKKYSLIVAHVEK-EGGKNNIQAQ-----QNLE--- 1832

Query: 681 NKYNYSQQVIQDLNKEIALQKESLMSLQALQDKAL--QKEKHYLQTTITKEAYDALSR-K 737
 N ++ Q+ +Q+ KE+ Q L ++ +LD L QKE H ++ Y+ L +
 Sbjct: 1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVRQKEVHRVEMEELTSKYEKLOALQ 1888

Query: 738 SAACQDDLTQALEKLNHVTSETKSLQQSLTQTQEKKAQ-LEEEIAYEERMKKLNTL-- 794
 ++ T+ LE+ S++ +Q L E + LE ++ E +KL E+
 Sbjct: 1889 QMDGRNKPTELLEENTEEKSKSHLVQPKLLSNMEAQHNDEFLKLAGAEREKQKLGKEIVR 1948

Query: 795 --RKLGRGFHQESELEVHAFDCKLEEMSCQVLQWQKHQNDLKMALAAKEEQLEFQEEA 852
 + LR +E + E+ K+ ++ + ++ Q+Q +LK + ++ +REF ++A
 Sbjct: 1949 LQKDLRMLRKEHQEILEIKKEYDQEREKIK-QEEDLELKHNSLTKQLMREFNTQLAQ 2007

Query: 853 LKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLREQV 912

++ L KE Q V + + Q TN Q K K+A EK + R
 Sbjct: 2008 KEQELEMTIKETINKAQ-EVEAEELLES-----QEETN--QLLK--KIA-EKDDDLKRTAK 2057

Query: 913 NYIAKLSGEKDHLSVMVHLQENKKLKEIEEKKMAEN 952
 Y L ++ + + + LQ + ++L+K+ ++K + EN
 Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKYYQQKLEQEEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22
 Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLSNLSLQLDIKN--LHDVCKRQRKTLDNQLCMEEAMNSSHDKK 58
 +KD+ + +N K L +LD+K L + + L+ +EE ++ D+
 Sbjct: 657 LKDKKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEEELSVLKDQT 714

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLV-LQEELEFHTTELQTSYYSLRQYQSI 117
 +E E + K H +Q+ + K+ V +Q+ + +++ L++
 Sbjct: 715 DKMK---QELEAKMDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKRDKH 771

Query: 118 LEKQTSDLVLLHHCKLKEDEVILYEEEMG---NHNENTGEKLHLAQEQALAGDKIASL 174
 L++ + + L K E E+ ++ ++ T E+ +EQLA K+ L
 Sbjct: 772 LKEHQAHVENLEADIKRSEGELQOASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDL 831

Query: 175 ERSNLNLYRDKYQSSLSNIELLECQVKMLQELGGIMQO-EPENKGDHVKVRIYTPCMIO 233
 E L + + + + + + + + +M Q E +N KV+ T
 Sbjct: 832 ETERILLTKQVAEVEAQKQDVCTELDAHKIQVQDLMLQLEKQNSEMEQKVKSLTQ-VYES 890

Query: 234 EHQETQKRLSEVWQKVSQQDDLIQELRN---KLACSNALVLEREKALIKLQADFASCTA 289
 + ++ K + Q + ++++I ++R ++ + +E ++ L ++ +
 Sbjct: 891 KLEDGNKEQETKQILVEKENMILQREGQKKEIILTQKLSAKEDSIHILNEEYET--- 947

Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349
 ++ + ++ E +K+ K +QE + L E L K+L +S++++
 Sbjct: 948 --KFK-NQEKMEKVKQKAKEMQETLKKLLDQEA--KLKKELENTALELSQKEKQFNAK 1002

Query: 350 MMKL-ELDLHGLREETSA-HIERKDKDITILQCRQLQELQLEFETQKLTLLKDKFLQEKD 407
 M+++ + + G+ + S +K++ ++ + +EL + +K ++ + LQE
 Sbjct: 1003 MLEMAQANSAGISDAVSRLTNQKEQIESLTVHREELNDVISIWEKKLNQAAEELQEIH 1062

Query: 408 EM-LQELEKKLTQVQNSLLK---KEKELEKQQCMATE----LEMTVKEAKQD-KSKEAEC 458
 E+ LQE E+++ ++ +L +++E+ K+ E + T+ E ++ K K A
 Sbjct: 1063 EIQLEKEQEVAELKQKILLFGCEKEEMKEITWLKEEGVKQDITLNELEQLKQKSAHV 1122

Query: 459 KALQAEVQKLKNSLEAKQERLAAQQAQKKEAALAGCHLEDTQKRLQKGLLLDKQKA 518
 +L + KKK LE+ + + ++ +E+ E+ +RK+ + L K K
 Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNLKSLKENTFLQEQVLVELKMLAEEDKRVSE--LTSKLT 1180

Query: 519 DTIQELQRELOMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKEKRLQKTVA 578
 T +E Q +K + E + +K EEL+++L +K E + K + +
 Sbjct: 1181 -TDEEFQSLKSSHEKSNKSLKEDKSLFKKLEELAIQLDICKKTEALLEAKTN--ELIN 1237

Query: 579 EQDMKMNDMLDRIKH-QHREQGSIKCKLEEDLQEQATKLEDKREQLKKSKEKLMGEL 637
 K N +L RI H QHR K+E L T + + QL++ E + +
 Sbjct: 1238 ISSSKTNAISLRISHCQHRTT-----KVKEALLIKTCTVSELAQLRQLTBEQNTLNISF 1292

Query: 638 EALRQEFKKKD---KTLKENSRLKEENENLR-----AELQCCSTQLESSL---- 680
 + + ++K+ K++K + L E E L+ +E + C TQL+ L
 Sbjct: 1293 QQATHLEKENQIKSMKADIESLVTEKEALQKEGGNQQQAASEKESCITQLKKELSENI 1352

Query: 681 NKYNSTQQVIQDLNKEIALQKESLMSLQALQKALQ-KEKHYLQTTITKEAYDALSRKSA 739
 N ++ +++ EI+ + L L QL ++ EK +++ K+ YD +
 Sbjct: 1353 NAVTLMKEELKEKKVEISSLSKQLTDNLVQLQNSISLSEKAAISSLRKQ-YDEEKCELL 1411

Query: 740 ACQDDLTQALEKLN-HVTSETKSLQQLTQTQEKKAQLEEEIIAYEERMKKLNTLRL-KL 797
 DL+ ++ L+ S + + + E K + + ++ +K+L +L K
 Sbjct: 1412 DQVQLDSFKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELKS 1471

Query: 798 RGFHQESELEVHAFDCKLEEMSCQVLOWQKQHNDLKMALAAKEEQLR-EFQEEMAALKEN 856
 + +++ E +++ ++L++ + + + ++D + KE L E + + A + E
 Sbjct: 1472 KEAYEKDE-QINLLKEELDQONKRFDCLEKEMEDDKSKMEKKESNLETELKSTARIME- 1529

Query: 857 LLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIA 916
 LED + + T + N+ ++ N Q QK K +L +++ +
 Sbjct: 1530 -LEDH-----ITQKTIEIESLNE-VLKNNYNQ----QKDIEHK---ELVQLQHFQ 1570

Query: 917 KLSGEKDH----LHSMVHLQENKKLKEIEEKKMAENTRLCTKA 959
 +L EKD+ ++ L+ + +K E+E KK + E+ L K+
 Sbjct: 1571 ELGEEKDNRVKEAEKILTLENQVYSMAELETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22
 Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMNSSHDKKQAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQQLQEFHTEELQT 106
 + E N + + Q EE E + S K ++ L + LQ+E +
 Sbjct: 1281 LTEEQNTLNISFQQATHQLEEKENQIKSMKA----DIESLVTEKEALQKEGGNQQQAASE 1336

Query: 107 SYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNENTGEKHLAQEQALAL 166
 + Q + L + + + L+ K K+ E+ +++ + N + L++++ A
 Sbjct: 1337 KESCITQLKKELSENINAVTLMKEELKEKVEISSLSQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGELGGIMGQEPENKGDHKSVMRIY 226
 I+SL + Y ++ L ++ L +V L E + Q + S+ +
 Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVWQKVSQQDDLIQEL--RNK-LACSNALVLE--- 272
 + HQ T K L E ++K Q + L +EL +NK C + +
 Sbjct: 1448 KAQSRTQHONTVKELQIQLELKSKEAYEKDEQINLLKEELDQKNRFDCLKGEMEDDKS 1507

Query: 273 -REKALIKLQADFASCTAT----HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQN 327
 EK L+ + S TA + + E E + ++LK+ +QKD E++
 Sbjct: 1508 KMEKKESNLETELKSTARIMELEDHITQKTIEIESLNEVLKNYNQKDI-----EHKE 1561

Query: 328 LVKDLRVELEAVSEQKRNIMKMDMMKLELDLHGLREETSATIERKDKDI--TILQCRLOEL 385
 LV+ L+ + + E+K N +K+ + L L A +E K K++ L + +E
 Sbjct: 1562 LVQKQLQ-HFQELGEEKDNRVKEAEKILTLENQVYSMAKELETKKKELEHVNLSVKSKEE 1620

Query: 386 QLEFTETQKLTLLKDKFLQEKDEMLQLEKRLTQVQNSLLKKEKELEKQQCMATELEMTV 445
 +L+ E + L+ + + E+ ++ E+K+ ++ LL + +E E+Q TE ++
 Sbjct: 1621 ELKALEDR---LESES-AAKLAELKRKAQKIAAIAKKQLLSQMEKEEQYKKGTESHLSE 1676

Query: 446 KEAKQDKSKEAECKALQAEVQKLKNSLEEAKQERLAAQQAQCK-EEAALAGCHLEDTO 504
 K + +E E L+ +++ ++S E RA AA + EEA GC + +
 Sbjct: 1677 LNTKLQE-REREVHILEELKLSVESSQSETLIVPRSAKNVAAYTEQEEADSQGCVCQKTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLE 564
 K+ +L + + + LQR Q +KE +++ + R + +E ++L A K
 Sbjct: 1736 EKIS---VLQRNLTEKEKLLQRVGQ--EKEETVSSHFE--RCQYQERLIKLEHAEAKQH 1788

Query: 565 NSDKEKRQLQKTVAEQDMKMDMLDRIKHQHQREQ--SIKCK--LE---EDLQ-----E 611
 LQ+ + E++ K + ++ +H +E G +I+ K LE +D+Q E
 Sbjct: 1789 EDQSMIGHLQEELEEKNNKYSLIV--AQHVEKEGGKNNIQAQONLENVFDVQKTLQKEKE 1846

Query: 612 AT-KLEDEKREQLKKSKEHEKLMEG-ELEALRQEFKKDKTLKENS---KLEENENL 665
 T ++LE K ++L +K + E+E L +++K + + R +L EEN
 Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQQMDGRNKPTELLEENTEE 1906

Query: 666 RAELQCCSTQLESSLN-KYNTSQQVIQDLNKEIALQKESLSLQAQLDKALQKEKHLYQT 724
 +++ +L S++ ++N + + +E + ++ LQ L + L+KE H +
 Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDL-RMLRKE-HQEL 1964

Query: 725 TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQSLTQTQEKKAQLEEEI IAYE 784
 I K+ YD R+ Q+ + LE L H ++ + +++ TQ +K+ +LE I +
 Sbjct: 1965 EILKKEYDQ-EREERIKQEQ--EDLE-LKHNSTLKQLMREFNTQLAQKEQELEMTI---K 2017

Query: 785 ERMKKLNTELRLKRGFHOESELEVHAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLR 844
 E + K +L HQE E + KK+ E + + K+++ ++L A+EE++
 Sbjct: 2018 ETINKAQEVEAELLESHEQE---ETNQLLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEAALKENLLEDDKEPCCLPQWSVP-KDTCRLYRGNDQIMTNLEQWAKQKQVANEK 903
 ++ E L + ++ L Q P D + ++ T L Q K +++ K
 Sbjct: 2072 AKVRDLQTQLEELQKKYQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTTLISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDLHLSVMV-HL 932
 L Q REQ++ + +L + +++++ V HL
 Sbjct: 2124 LKEQEFREQIHNLEDRLKKYEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20
 Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQLDIKN-LHDVC-KRQRKTLQDNQLCMEEAM-NSSHDKK 58
 ++ ++ +E+ +L KLL + +K L + + +K Q N +E A NS+
 Sbjct: 957 EKVKQKAKEMQETLKKKLLDQEAQLKKELENTALELSQKEKQFNAKMLEMAQANSAGISD 1016

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQQLQEFHTEELQTSYSLRQYQSIL 118
 L + E + S + H R+L + + + ++L EELQ + ++ +
 Sbjct: 1017 AVSRLNETNQKE-QIESLTEVHRRELNDV---ISIWKKLNQQAELQ-EIHEIQLEK-- 1069

Query: 119 EKQTSIDLV--LLHHHCKLKE-DEVILYEEEMGNHNENTGEKHLAQEQALAGDKIASLE 175
 E++ ++L +L C+ +E ++ I + +E G + T +L +Q + + +A E
 Sbjct: 1070 EQEVALKQKILLFGCEKEEMNKETITWLKEEGVKQDTTLNELQEQKQSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVKMLQGELGGI--MGQEPENKGDHKSVMRIYTPCMIQ 233
 L + +K + L N L E LQ +L + + +E + K ++ T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQLVELKMLAEEDKRRVSELTSLKKTDEEFQ 1186

Query: 234 E----HQETQKRLSEVWQKVSQQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNKSLLED---KSLEFKKLSEELAIQLDICKKTEALLEAKTNELINISSST 1243

Query: 286 SCTATH-RYPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKR 344
+ + + + + + + + I + + + Q + E QN + + E+K

Sbjct: 1244 NAILSRISHCQHRTTKVKEALLIKTCTVSELAQLRQLTEEQNTLNISFQQATHQLEEKE 1303

Query: 345 NIMKDMKLELD-LHGLREETSABIERKDKDITILQCRLQELQLEFTET-QKLTLLKDKF 402
N + K M K +++ L +E + + + + + +L+ E +E +TL K++

Sbjct: 1304 NQIKSM-KADIESLVTEKEALQKEGGNQQAASEKESCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQEKDEMLQLEKRLTQVONSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKAL 462
L+EK + L K+LT + N L+ L +++ + L E K + + + L

Sbjct: 1362 LKEKKEISSLKQLTDL-NVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQ--DLS 1418

Query: 463 AEVQKLKNSLEEAKQERLAAQQAACKEEAALAGCHLEDTQRKLQKGLLLDKQA---- 518
+V L A +Q + + + + K++ +T +LQ L L ++A

Sbjct: 1419 FKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELKSKEAYEKD 1478

Query: 519 DTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKEKRQLQKTVA 578
+ I L+ EL K + E ++ ++E+ L +L++ +L+ +

Sbjct: 1479 EQINLLKEELDQKNRFDCLKGEMDDKSKMEKKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMDMLDRIKHQHREQGSICK-LEEDLQEATKLEDDKREQLKKSKEHEKLMGEL 637
++ +++ + + +K+ + +Q I+ K L + LQ +L E+K ++K+++ +E ++

Sbjct: 1536 QKTIEIESLNEVLKN-YNQKQDIEHKELVQKLQHFQELGEEKDNRVKEAEEKILTLENQV 1594

Query: 638 EALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLES-SLNKYNTSQQVIQDLNKE 696
+++ E + K K L+ + + + E L+A L+ +LES S K ++ + ++

Sbjct: 1595 YSMKAELETKKKELEHVNLSVKSKEELKA-LE---DRLESESAAKL---AELKRKAEQK 1647

Query: 697 IALQKESLMSLQALDKALQKEKHLYQTITKEAYDALSRKSAACQDDLTQALEKLNHVT 756
IA K+ L+S Q++ +KE+ Y + T + L+ K + ++ EKL V

Sbjct: 1648 IAAIKKQLLS---QME---EKEEQYKKG--ESHLSELNTKLQEREREVHILEEKLKSV 1699

Query: 757 S---ET----KSLQQSLTQTEKKAQLEEEII-AYEERMKKLNTLRLKRGFHEQSELEV 808
S ET +S + T++++A + + YEE++ L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEEADSGQCVQKTYEEKISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLEFQEEAALKENLLEDDKEPCCLP 868
++ EE + + Q+Q L L E + E Q + L+E L E +K+ +

Sbjct: 1753 QRVGQKEETVSSHFMRCQYQERLIKLEHAEAKQHEQDSMIGHLQEELEENKKYSLIV 1812

Query: 869 QWSVPKDCRLYRGNDQIMTNLEQ-WAKQKQVANER-LGNQLREQ-VNYIAKLSGEKDHL 925
V K+ + N Q NLE + QK EK L Q+ EQ + + + +

Sbjct: 1813 AQHVEKEGGK---NNIQAKQNLNVFDDVQKTLQEKELTCQILEQIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQQENKKLK 940
H V M L + +KL+

Sbjct: 1870 HRVEMEELTSKYELQ 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHOETQKRLSEVWQKVSQQDDLIQE-LRNKLACSNAIV-LEREKALIKL-QADFASCTA 289
+E +TQ ++ +V + L + ++ L S++ L R + L + D S TA

Sbjct: 53 RESGDTQSFAQKLQLRVPSVESLFRSPIKESLFRSSKESLVRTSSRESLNRDLDSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNMKD 349
+ P E ED+ L +++ Q L + + R + + + + +

Sbjct: 113 SFDPPSDMDSEADLVGNSDSLNEQLTQRLR--RMERSLSSYRGKYSSELVTAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSABIERKDKDIT-ILQCRLQELQLEFTETQKLTLLKDKFLQEKDE 408
KL+ G+ ++ +DK + I + R +ELQ++ + L + D L+EKD+

Sbjct: 171 KKKLQ----GILSQS-----QDKSLRRIAEELR-EELQMDQQAQKHLQEEFDASLEEKDQ 219

Query: 409 MLQELEKRLTQVONSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQAE---V 465
+ L+ +++ ++ L ++ + + +LE + + + + + E++ + + + V

Sbjct: 220 YISVLQTVQSLLKQRLRNGPMNVDLKPLP-QLEPQAEVFTKEENPESDGEPPVEDGTSV 278

Query: 466 QKLKNSLEEAKQERLAAQQAAC-KEEAALAGCHLEDTQRKLQKGLL-LDKQKADTI 521
+ L+ + K+QE L ++ Q KE+ L E Q +L + L L+K K +

Sbjct: 279 KTLTQQRVQRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQDLERLQELEKIKDLHM 338

Query: 522 ELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKEKRQLQKTVAEQD 581
E + + L+ ++ E+ + + E ++ E L E + R K + Q

Sbjct: 339 AEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRMHETLEMKEEEIAQLRSRIKQMTTQ 398

Query: 582 MKMNDMLDRIKHQHREQGSIKCKLEEDLQEAT-KLLEDKREQLK---KSKEHEKL-MEGE 636
 ++ + ++ + E+ + +EA KL + EQ+K K+ E E++ ++ E
 Sbjet: 399 EELREQKESERAAFELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERISLQOE 458

Query: 637 LEALRQEFKK-KDKTLKENSRRKLEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNK 695
 L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K
 Sbjet: 459 LSRVKQEVVDVMKKSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQALEKLN-H 754
 +AL+K L+ +K Q+ + + K+A S DL Q E
 Sbjet: 513 -VALEKSQSEYLLKISQEKEQESLALAELELQKKAILTESENKLR---DLQEAETYRTR 568

Query: 755 VTSETKSLQOQSLTQTEKKAQLEEEIIAYEERMKKLNTLRLRGFHOESELEV--HAFD 812
 + SL++SL QE K Q ++ + E K N E+ + H+ +ELE H D
 Sbjet: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHNKEITVMVEKHK-TELESLEKHQD 624

Query: 813 KLEEMSCQVLQWQKHQNDLKLAAKEEQLE---FQEMAALKENLLED-DK 862
 E QVL+ +Q+Q +++ L K EQ +E FQ + + E LE D
 Sbjet: 625 ALWTE-KLQVLK---QQYQTEMEKLEKCEQEKETLLKDKETIFQAHIEEMNEKTLEKLDV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANKEKLGNOLEQVNYIAKLSGEK 922
 + L S+ + + + ++ L Q ++L ++ EQ N+ +
 Sbjet: 682 KQTELE---SLSELSEVLKARHKEEELSVLKDQTDKMKQELEAKMDEQKNHHQQQVDSI 739

Query: 923 DHLHVMVHLQQENKLLKKEIEEKKM 948
 H V + Q+ K LK +I + ++
 Sbjet: 740 IKEHEVSI--QRTKALKDQINQLEL 763

Score = 193 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09
 Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQELEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467
 M ++L++K+++ Q L + + +T M + + ++ E + Q
 Sbjet: 1 MFKKLKQKISEEQQLQALAPAQASSNSSTPTRMRSRTSFTQQLDEGTPNRESGDTQS 60

Query: 468 LKNSLE-EAKQERLAAQQAQCKEEAALAGCHLEDTQKRLQGLLLDKQKA--DTIQEL 524
 L+ E L + ++ + + R+ L LD A D ++
 Sbjet: 61 FAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLODOSTASFDPPSDM 120

Query: 525 QRELQMLQKESSMAEKEQTSNRKRVEELSL---ELSEALRKLENSQKEKRLQKTVAE 579
 E + L S KEQ R R E SL + SE + + +EKK++LQ +++
 Sbjet: 121 DSEAEDLVGNSDSLNEQLIQLRRMERSLSSYRGKYSSELVTAYQMLQREKKKLQGISQ 180

Query: 580 -QDKMNDMLDRIKHQHREQGSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632
 QD + + + + +Q+ K EE L+E + +L+ + LK+ + +
 Sbjet: 181 SQDKSLRRIAELEELQMDQAKKHLQEEFDASLEEKDQYISVLQTVSLLKQRLRNGPM 240

Query: 633 MEGELEALRQ-EFKKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQ 688
 L+ L Q E + + T +EN E E+ L+ +++ N ++
 Sbjet: 241 NVDVLKPLPQLEPQAEVFTKEENPESDGEPPVEDGTSVKTLETQQRVKKRQENLLKRCKE 300

Query: 689 VIQDLNKEIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQA 748
 IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +
 Sbjet: 301 TIQSHKEQCTLLTSEKEALQEQLDERLQ-ELEKIKDLHMAEKTCLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQOQSLTQTEKKAQLEEEIIAYEERMKKLNTLRLRGFHOESELE 807
 LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E
 Sbjet: 358 LEQDKGMVIAETK---RQMHELTLEMK---EEEIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKHQNDLKLAAKEEQLEFQ---EEMAALKENLLEDDKE 863
 AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E
 Sbjet: 410 RAAF---EELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERISLQOELSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKQVANKEKLGNOLEQVNYIAK 917
 + + S + +L + +++ + EQ K+ + + Q++ Q Y+ K
 Sbjet: 466 VVDVMKKSSEEQIAKLQKLHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDHLHVMVH-LQQENKLLKKEIEEK---KMKAEENTRLCTKALGPSRTESTQREK 972
 +S EK+ S+ + L+ + K+ E E K + +AE R L S +S Q K
 Sbjet: 525 ISQEKEQESLALAELELQKKAILTESENKLRDLQEAETYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3_lg13, frame 1

Report for DKFZphtes3_lg13.1

[LENGTH] 1007
 [MW] 117480.77
 [pi] 5.90

[HOMOL] TREMBL:AF092090_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
5e-15
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08
[FUNCAT] 1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1322] 4e-06
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w]
MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 -
myosin-1 isoform] 3e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04
[EC] 3.6.1.32 Myosin ATPase 1e-16
[PIRKW] nucleus 3e-10
[PIRKW] phosphotransferase 6e-09
[PIRKW] duplication 2e-06
[PIRKW] citrulline 2e-12
[PIRKW] tandem repeat 1e-16
[PIRKW] endocytosis 2e-13
[PIRKW] heart 8e-13
[PIRKW] transmembrane protein 1e-13
[PIRKW] serine/threonine-specific protein kinase 6e-09
[PIRKW] zinc finger 2e-13
[PIRKW] metal binding 2e-13
[PIRKW] DNA binding 4e-12
[PIRKW] muscle contraction 1e-16
[PIRKW] acetylated amino end 1e-11
[PIRKW] actin binding 1e-16
[PIRKW] mitosis 5e-15
[PIRKW] microtubule binding 5e-15
[PIRKW] ATP 1e-16
[PIRKW] thick filament 1e-16
[PIRKW] phosphoprotein 4e-16
[PIRKW] skeletal muscle 2e-14
[PIRKW] calcium binding 2e-12
[PIRKW] alternative splicing 1e-16
[PIRKW] coiled coil 1e-16
[PIRKW] P-loop 1e-16
[PIRKW] heptad repeat 3e-10
[PIRKW] methylated amino acid 1e-16
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 2e-13
[PIRKW] cardiac muscle 8e-13
[PIRKW] hydrolase 1e-16
[PIRKW] microtubule 3e-10
[PIRKW] muscle 8e-13
[PIRKW] EF hand 2e-12
[PIRKW] cytoskeleton 2e-15
[PIRKW] hair 2e-12
[PIRKW] calmodulin binding 2e-13
[PIRKW] Golgi apparatus 3e-10
[SUPFAM] myosin heavy chain 1e-16
[SUPFAM] conserved hypothetical P115 protein 1e-07
[SUPFAM] centromere protein E 5e-15
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09
[SUPFAM] calmodulin repeat homology 2e-12
[SUPFAM] myosin motor domain homology 1e-16
[SUPFAM] alpha-actinin actin-binding domain homology 2e-07
[SUPFAM] plectin 2e-07
[SUPFAM] trichohyalin 2e-12
[SUPFAM] pleckstrin repeat homology 8e-08
[SUPFAM] ribosomal protein S10 homology 2e-07
[SUPFAM] giantin 3e-13
[SUPFAM] protein kinase homology 6e-09
[SUPFAM] protein kinase C zinc-binding repeat homology 8e-08
[SUPFAM] kinesin motor domain homology 5e-15
[SUPFAM] human early endosome antigen 1 2e-13
[SUPFAM] M5 protein 1e-07
[PROSITE] LEUCINE_ZIPPER 7
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 20

[illegible]

Prosites for DKFZphtes3_lg13.1

(No Pfam data available for DKFZphtes3_1g13.1)

group: cell structure and motility

DKFZphtes3 lk11 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, compete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```

1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGGCGG GGATCGTCTC
101 CGGCACCTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGCG
151 GAGCAGCACG GGTCCATGA ACGTCACCC CTCCACAAG GCCTCCCACC
201 CGGACTGTGT GCTGGCCAC CTCAACACG TCGCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTGGGC GGGCGACCG GCCTTCCCTT GTCCCGTGC
301 CGTGCTGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTT AGCCATGGCC
351 TTCGGGAGAG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCCTCAC GCATCGCCAT
451 CAACGAGGAG AACGCTGAGT CACTGCTGGA GGCAGGCGAC ATGCTGCAGT
501 TCCACGATGT GCGGGATGCT GCCGCCGAGT TCCTGGAGAA GAACCTTTTC
551 CCCTCCAACT GCCTGGGCAT GATGTGCTC TCGGACGCCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCCT GGCGCATGTG CCTGGTGCAC TTTGAGACGG
651 TGAGGACAGAG CGAGGACTTC AACAGCCTGT CCAAGGACAC ACTGTGGAC
701 CTCATCTCGA GTGATGAGCT GGAGACCGAG GACGAGCGGG TGGTCTTCGA
751 GGCCATCCTC CAGTGGGTGA AGCACGACCT GGAGCCACGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGCAGC GTGCGTCTGG CCTTGCTGCC GTCCGACTGC
851 CTGACGAGAG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACCAGG ATCCTGCAGA
951 ATGATGGCGT GGTCAACAGC CCCTGTGCCC GGCCACGCAA GCGGGGCCAC
1001 ACGTACTTCA TCCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACCAC AAGGCCAAGG AGATCATCCC CAAGGCCGAC CTGCCACGCC
1101 CCCGGAAGGA GTTCAGCGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGTCTCC AAGGATGTCT GGGGTGACGA
1201 CACCGTACAT GAGGAATGGT CCAAGGCGGC GCCATGCTG ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAAGT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTCCCGGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGGCGGA TGGCGTCAGC AATGCCGCG TGGTGAGTGC CAAGCTGAAG
1451 CTCCTTTGTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAGTGCTAT GACCCCTCGG AGAACAGGTG GACGATCAAG GCCGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCCGCTGCCG TCCTGGGCAG CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTACAGCC GCCTCGGCCT ACCGCTTTGA
1651 CTGTGAGACC AACCAGTGGG CGCGGATTGG GGACATGACT GCCAAGCGCA
1701 TGTCTTGCCA TGCCCTGGCT TCCGGCAACA AGCTCTATGT GGTGGGGGGC
1751 TACTTTGGGA CCCAGAGGTG TAAGACTCTG GACTGCTATG ACCCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTGCC CTACTCACTT ATCCCCACGG
1851 CCTTTGTCTG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCCAGCCAGA CCGCGGCCCT CAGTGTACCA GCGTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGGTGGAAGG
2001 GGCCCTGCCA GCTCTGGGGA GCAGCAGCCT TGGGCTGTTC TGAGCTTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGCATCCG TGCCCTTGGG GGCTCTTCA
2101 GCTTTGCACT GGTTTGTGGG AAGACATACC TCCCAGAGGG GCATGGACTG
2151 CCACCAAGGAC TGACCTGGC GTCGGGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACCTG TTTGGCAGGT CCTGGACTGG GGCCGGGCAG GCAGGGGCAG
2251 GGAGGCGCCC CGGGTGGGCT TTGGGGCTGC GGCCTGCCA CACATCCTTT
2301 CCCTCCTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGTG
2351 CTGGGCTTGG GAAACTAGGT TCCAGGGGT TGAGACCAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAACTGCA GGGGGGCCCT AGTGACATCC
2451 ATGAGGCCTT ATTAGCAAAG GACACCCAGA CCTCCAAGGT TTGTGGGGCC
2501 CTCCACAAA GCTGTAAGTC CCAGCCACC TACTCAGGGC CTTGCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCTCCCTC
2651 CCCCTCTCA GAGCCACCC TGAGAGGCAG CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GCTTCTCCCC AGGAGACCAC GTCGCCCTCC

```



```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCCT CACCATCCTC TTGCGTTTGT GCCCCCATTT GCTCCCTGAG
2901 CTCACAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCAATTGC
2951 TGTCACCTCC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTTGGCG CCTGTCACTG ACCATCCTAA TGCCTCTCGC
3051 TCACTCCCAA GCCACCATT TTAGAGGGAGG GGTGTTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC GCGGAGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCCCTC TTGGGGGTGG GACCTTTGGG CAGGTTTTC CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCCTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTTGCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

98350113:

Cloning of human ENC-1 and evaluation of its expression and regulation in nervous system tumors.

97252647:

ENC-1: a novel mammalian kelch-related gene specifically expressed in the nervous system
encodes an actin-binding protein.

98234394:

NRP/B, a novel nuclear matrix protein, associates with p110(RB) and is involved in neuronal differentiation

Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589

Category: strong similarity to known protein

Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLT LR KHCMTDVTIL
51 WAGDRAFPCH RAVLAASSRY FEAMFSHGLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAYSSRI AINEENAESL LEAGDMLQFH DVRDAAAEFL EKNLFPSNCL
151 GMMLLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK DTLDDLISSD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEA
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGQTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVYTGGRGS
351 ENGVSXDVVW YDTVHEEWSK AAPMLIARFG HGSAELENC L YVVGHTSLA
401 GVFPASPSVS LKQVEKYDPG ANKMMWVAPL RDGVSNAAV SAKLKLFEVFG
451 GTSIHRDMVS KVQCYDPSEN RWTIKAECPO PWRYTAAAVL GSQIFIMGGD
501 TEFTAASAYR FDCETNQWTR IGDMTAKRMS CHALASGNKL YVVGGYFGTQ
551 RCKTLDCYDP TSDTWCNITT VPYSLIPTAF VSTWKHLPA

```

BLASTP hits

Entry MMU65079_1 from database TREMBL:

gene: "ENC-1"; product: "actin-binding protein"; Mus musculus

actin-binding protein (ENC-1) mRNA, complete cds.

Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611_1 from database TREMBLNEW:

gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens

nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.

Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314_1 from database TREMBL:

gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,

complete cds.

Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Alert BLASTP hits for DKFZphtes3_1k11, frame 2

No Alert BLASTP hits found

Pedant information for DKF2phtes3_1k11, frame 2

Report for DKFZphtes3_1k11.2

```

[LENGTH]          589
[MW]               65923.45
[pI]               6.10
[HOMOL]            TREMBL:MMU65079_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds. 0.0
[FUNCAT]           10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
2e-09
[BLOCKS]           BL01016D Glycoprotease family proteins
[PIRKW]            zinc finger 1e-08
[PIRKW]            DNA binding 1e-08
[PIRKW]            transcription factor 1e-08
[SUPFAM]           POZ domain homology 3e-68
[SUPFAM]           vaccinia virus 59K HindIII-C protein 1e-15
[SUPFAM]           A55R protein 5e-29
[SUPFAM]           hypothetical protein YHR158c 4e-08
[SUPFAM]           A55R protein middle region homology 5e-29
[SUPFAM]           myxoma virus M9-R protein 1e-14
[SUPFAM]           A55R protein carboxyl-terminal homology 5e-29
[KW]               Alpha Beta

```

```

SEQ      MSVSVHETRKSRSSTGSMNVTLFHKASHPCDVLHLNLTLRKHCMFTDVTLLWAGDRAFFCH
PRD      cccccccccccccccccceeeeeccccchhhhhhhhhhhhhhhheeeeeccccchhh

SEQ      RAVLAASSRYFAMFSHGLRESRDDTVNFQDNLHPEVLELLLDFAYSRIAINEENAESL
PRD      hccccccccccccccccchhhhhheeeeeccccchhhhhhhhhhhhhccceeehhhhhhh

SEQ      LEAGDMLQFHDVDRAAAEFLEKNLFPSCNCLGMLLSDAHQCRRLYEFSWRMCLVHFETVR
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      QSEDFNSLSKDTLLDLISSDELETEDERVVFEAILQVWKHOLEPRKVHLPPELLRSVRAL
PRD      hhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhc

SEQ      LPSDCLQEAVSSEALLMADERTKLIIMDEALRCKTRILQNDGVVTSPCARPRKAGHTLLIL
PRD      cchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccccccccccccccceeeee

SEQ      GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVVW
PRD      cccccccccceeeccccccccccccccccccccceeeeeceeeeeccccccccccccceeee

SEQ      YDVTVEHWSKAAPMLIARFHGHSAELENCLYVVGHTSLAGVFPASPVSLSLKQVEKYDPG
PRD      cccccccccccccccccccccceeeccceeecccccccccccccccccccccccccccccc

SEQ      ANKWMWVAPLRDGVSNAAVVSAKLKLFFVGGTSHRDMVSKVQCYDPSSENRTWIKAECPQ
PRD      cceeeccccccccccccceeeccccceeecccccccccccccccccccccccccccccccc

SEQ      PWRYTAAAVLGSQIFIMGGDTFTAAASAYRFDCETNQWTRIGDMTAKRMSCHALASGNKL
PRD      cccccceeeccccceeeccccccccccccccccccccccccceeeccccccccccccceeeccce

SEQ      YVVGCGYFGTQRCKTLDCYDPTSDTWCNITTVPYSLIPTAFVSTWKHLPA
PRD      eeeccccccccccccccccccccccccceeeccccccccceeecccccccc

```

(No Prosite data available for DKFZphtes3_1k11.2)

(No Pfam data available for DKFZphtes3_1k11.2)

DKFZphtes3_1n3

group: signal transduction

DKFZphtes3_1n3 encodes a novel 1196 amino acid protein with similarity to *S. pombe* Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```

1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAACTTG TCAGGTCTGA AGAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACACAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCCATATT AAAGAACTA
251 CAAGTGATGA TGTAAGTGCT GCTAACACTA ACAACCTGAA GAAGAGCACG
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAATAAAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAGGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGAAT TGATGCAAGC ATATCAGTGC
601 CATGTAACGT AAGAAATGGC AAAGGAGATT AAGAGGAAAA TAAGAAGAA
651 ACTGAAAGAA CAGTTGACTT ACTTCCCTC AGATACTTTA TTCCATGATG
701 ACAAACTAAG CAGTGAAAAA AGGAAAAAGA AAAAGGAAGT TCCAGTCTTC
751 TCTAAAGCTG AAACAAGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
801 TGAACAAAAG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGCTCA ATGGAACAAA GCACAGAAGA CAGCATGCAA
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAAGAAG CTAAGGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTTCA GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTACCCG AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGTGGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTTCTCTTA CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTTGCTT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCCGAAAATT GCCTGGGCAT
1451 TTCTTAAAGT TCTGGGAGCC AATGGAAATG CAAACATCAA CTCAAAACCTT
1501 CGCTTGCAAG TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAAGTGT
1551 TGTGAGGCA TTTGAATGGT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACCTGA AGAGGACTGA AAGTTCCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAAG GTAAACCAGT
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGG AACGACTCCC TGGGCAGGCT
1801 TGCCGTATCC CAAACAAACA CCTCTTCTCA CTAATGTCAG GAGAACGAGG
1851 ATGTTTTTGT CTTGATTCTT CCCACAATGG AAGAATATTA GCAGCAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTATTT TATATGAAAT TCCTTCTGGA
1951 CGTTTCATGA GAGAATTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGGTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACTTT CAGAGTTTAA
2101 CCTCATCTTT CTTTGTGTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTTGATACT GAAGTTCATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGA ACATTCAAGT CACCACTGGA CTATAAATAA GGAAATTA
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTTGGAGA TTCTCCCAA
2451 TGGAAAACGT TTGTTAATCC ATACCAAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTC ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTTGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGA CCCAGAAACA GGAGAACAAG

```



```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTG AGACATTCT
2701 TATCATCCAT TTGAAATAT GGTGTCATTC TGTGCATTG GGCAAAATGA
2751 GCCAATTCTT CTGTATATTT ACGATTTCCTA TGTGCCCCAG CAGGAGGCTG
2801 AAATGTTCAA ACGCTACAAT GGAACATTTC CATTACCTGG AATACACCAA
2851 AGTCAAGATG CCCTATGTAC CTGTCCAAAA CTACCCCATC AAGGCTCTTT
2901 TCAGATTGAT GAATTTGTCC AACTGTAAAG TTCTTCAACG AAGATGCAGC
2951 TAGTAAACAA GAGGCTTGAA ACTGTACAG AGGTGATACG TTCCTGTGCT
3001 GCAAAAGTCA ACAAAAATCT CTCATTTACT TCACCACCAG CAGTTTCCTC
3051 ACAACAGTCT AAGTTAAAGC AGTCAAAACAT GCTGACCGCT CAAGAGATTTC
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAAGAAAG
3151 CCTTGTAACC ATCAGGTAGA TACAGCACC ACGGTAGTGG CTCTTTATGA
3201 CTACACAGCG AATCGATCAG ATGAACTAAC CATCCATCGC GGAGACATTA
3251 TCCGAGTGTT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAAACACT
3351 GTATCAAGAA CTGCCTCCTG AGATAAAGGA GCGATCCCCT CCTTTAAGCC
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA
3451 ATCAATAAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC
3501 ACATTCTGAA ATGAGAAAAG AACAGAGCCA TGAGGACCAA GGACACATAA
3551 TGGATACACG GATGAGGAAG AACAGCAAG CAGGCAGAAA AGTCACTCTA
3601 ATAGAGTAAA GAATTGAAGA AAAGTTAAGA GCTGCCGAAA TGCACAGAGG
3651 TGAATGATGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCAG
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTTGTCTT ATGAATGACT
3751 CTAGAAAAAT CAGAAATCAAG TTGTGGGTGG AAAAAATCAAC GTGGCCTTTG
3801 AGTTCAGTTG TTATAAACCA TTGTGACTAT TGTGTGCTAA AGTATTGGTA
3851 CTTATATTGT TAGTAATTGC ATCATAATTA CATTACCAGT GTTGGAAAAC
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAAAATGTA ATAAAAGGTG
3951 TTTGCGTTAT TAGGATGTCT GTTAAGTAAT CATTTAATAT TATTATATTG
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT
4051 TGAAATTACAC TTTATTTAAA AGATAAGCAG CTGACTGGGC ACGGTGCCTC
4101 ATGCTGTGTA TCCTAGCACC TTGGGAGGCT GAGGCAGGTG GATCACCTAA
4151 GGTGAGGACT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT
4201 CTACTAAAAA TACAAAAATC AGCCGGGTCT CATGGCAGGC ACCTGTAATC
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTCGAG
4301 TGAGCCAAGA TCACGCCATT GCACTCCAGC CTGGGGGACA GAGCAAGACT
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTTAGAATA TGGCGCATTC
4401 AAAACAGTCT CAGTAACAAA GACATTAAAA GAAAACAATT TACTTTCTAA
4451 TTAATAATTT GTGTTTCTTA AGATCAAAATC ATATAGGTAA CTTTCATAGC
4501 CTAATTTAAA AGTGATTTT GGCTGGACTG GCAACATGT TCCCAATGTC
4551 TTTACTTTTT AAAAAGGCTT TTTCATATTT AAGCACATAC CTATTTTGTA
4601 GACTTACATT GTTTAATATT TATTTAATC TTAATATTT TACATTATTA
4651 TATTGCATTA TTTATTTTT CTAAGTTCCA GAATAATAGT GTCATTATTA
4701 TAGACTATAT GTTTGAAGT TTGATATTAT AATGGGATAT TCATTTTTTG
4751 TTCTTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATAAAA
4801 TATTTAACCC CAAGAAAGTG AAAACTAATA TAAATTAGA AAGACCTATC
4851 CAAATTAGAC AGTCAATTCC ATTAAATAA GAAGTGAGAA AAACAATGTT
4901 GGGCATTGAG GTGTAAATTT TGCCAGATG TATACCCAGT GTGAAATATC
4951 TTCTAATAAA AATATATTT GCTCTTATCC CTGCACATGT AGAGGCATAA
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAA AAGGCATTTT
5051 TGAAAGTGTT GACTGGCACT GATAACTGGT GAAGCCTACA GCCATCCGCC
5101 CAAAAGTCTG TTCTGATGGC ACTGAGTTTT CATTGTTCTG ATGTATTAAG
5151 TCTGTGTGTC AGGTACAGCT GGGCCAGCC AGCTTGAGTC ACTCTGTAC
5201 AAGCTTGTGT TTTCTGTCT TGTGAATGCA CTTGATAATT TAAAAATAA
5251 AATATCTGTT TCTCTGCAA AAAAAA

```

BLAST Results

Entry HS32B1 from database EMBL:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 32B1
 Score = 4445, P = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:
 Human exon-trapped sequence from 6q24.
 Score = 965, P = 4.0e-35, identities = 193/193

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196
 Category: similarity to known protein


```

1 MPTAESEAKV KTKVRFEKLL KTHSDLMREK KKLKKKLVR EENISPDITR
51 SNLHYMKETT SDDPDTIRSN LPHIKETTSO DVSAANTNNL KKSTRVTKNK
101 LRNTQLATEN PNGDASVEED KQGKPNKKVI KTVPLTTQD LKPETPENKV
151 DSTHQKTHTK PQPGVDHQKS EKANEGREET DLEEDDELMQ AYQCHVTEEM
201 AKEIKRKRK KLKEQLTYFP SOTLFHDDKL SSEKRRKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHQDD EISSMEQSTE DSMQDDTKPK
301 PKTKKKTKA VADNNEDVDG DGVHEITSRD SPVYPRCLLD DDLVLGVYIH
351 RTDRLKSDFM ISHPMVKIHV VDEHTGQYVK KDDSGRPVSS YYEKENVDYI
401 LPIMTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDS PKVILFFEIL
451 DFLSVDEIKN NSEVQNQECG FRKIAWAFK LLGANGNANI NSKLRLQLYY
501 PPTKPRSPLS VVEAFEWWSK CPRNHYPSTL YTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCEHHHESS SVDTEPGLEE SKEVIKWKRL PGQACRIPNK
601 HLFSLNAGER GCFCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
651 CGHNLIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
701 YTAKFHFAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CFDTGEGHMY SGDCTGVIVV WNTYVKINDL EHSVHHWTIN KEIKETEFKG
801 IPISYLEIHP NGKRLLIHTK DSTLRIMDLR ILVARKFVGA ANYREKIHT
851 LTPCGTFLFA GSEDGIVYV NPETGEQVAM YSCLPFKSPI RDISYHPFEN
901 MVAFAFAGQN EPILLYIYDF HVAQQAEMF KRYNGTFPLP GIHQSQDALC
951 TCPKLPHQGS FOIDEFVHTE SSSTKMQLVK QRLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQSKLKQSN MLTAQEILHQ FGFTQTGIIS IERKPCNHQV
1051 DTAPTIVVALY DYTANRSDLE TIHRGDIIIV FFKDNEDWWY GSIQKGQEGY
1101 FPANHVASET LYQELPPEIK ERSPLSPPEE KTKIEKSPAP QKQSKINKKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMRKNKQAGR KVTILIE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_in3, frame 1

TREMBL:U92792_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces
pombe general transcriptional repressor Tup1 (tup1) mRNA, complete
cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible
35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa
protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin";
S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N =
2, Score = 228, P = 1e-13

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible
35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa
protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin";
S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CERO3E1_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1,
N = 1, Score = 215, P = 2.3e-13

SWISSPROT:YZLL_CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING
PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6
kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein
(Pmc733) mRNA, complete cds.
Length = 321

HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18
Identities = 59/225 (26%), Positives = 111/225 (49%)

Query: 647 MRELCCGLNIIYDLSWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFH 706
+ E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +F+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD--LLSASMDKTVRLW--QVGRDSCLKVSHTNYVTVQFQ 119

Query: 707 PAVRELVTGTCYDSMIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTGEGHMYSGDCTG 766
P +TGC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITGCTDGLVRIWDVRK----CLVVDWANSKEIVTAVCYRPDGGKAVAGTITG 174

Query: 767 VIVVWNTYVKINDLEHSHVHWTINKEIKETEFKGIPISYLEIHPNGKRLLIHTKDLTRI 826
++ +LE V ++N K + + Y P K+L++ + D+ +RI

Sbjct: 175 NCRYDASENRLELESQV---SLNGRKKSLHKRIVGFQYCPSDP--KKLMVTSQDAQVRI 229
 Query: 827 MDLRILVARKEVGAANYREKIHSTLTGCTFLFAGSEGDGIVYVWN 871
 +D +++ + G + ++ + TP G + + S+D +Y+WN
 Sbjct: 230 LDGAHVISN-YKGLQS-SSQVARSFTPDGDHIVSADDSRIYMWV 272

Pedant information for DKFZphtes3_ln3, frame 1

Report for DKFZphtes3_ln3.1

[LENGTH] 1196
 [MW] 137114.70
 [PI] 6.79
 [HOMOL] SWISSPROT:YKY4_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN
 C14B1.4 IN CHROMOSOME III. 8e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
 TAF90 - TFIID subunit] 4e-10
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit]
 4e-10
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c]
 9e-08
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YDL145c] 9e-08
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YMR116c] 4e-06
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w]
 4e-05
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05
 [BLOCKS] BL00024H
 [SCOP] dltbgsd_ 2.46.3.1.1 betal-subunit of the signal-transducing 3e-91
 [SCOP] dlqfc_ 2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14
 [SCOP] dlfmk_ 2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15
 [SCOP] dladsbl_ 2.21.2.1.7 (1-63) Hemopoietic cell kinase Hck [human (Hom 3e-15
 [SCOP] dlckal_ 2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13
 [SCOP] dlqwea_ 2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15
 [SCOP] dlshg_ 2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13
 [SCOP] dlprmc_ 2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15
 [SCOP] dlhsq_ 2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13
 [SCOP] dlabo_ 2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13
 [SCOP] dldefna_ 2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15
 [SCOP] dlsema_ 2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13
 [SCOP] dlgbqa_ 2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16
 [SCOP] dlckaa_ 2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15
 [EC] 3.1.4.3 Phospholipase C 2e-07
 [EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07
 [EC] 3.6.1.32 Myosin ATPase 7e-07
 [EC] 2.7.1.112 Protein-tyrosine kinase 8e-06
 [PIRKW] nucleus 2e-08
 [PIRKW] phosphotransferase 8e-06
 [PIRKW] plasma 4e-07
 [PIRKW] duplication 4e-07
 [PIRKW] phosphoric diester hydrolase 2e-07
 [PIRKW] tandem repeat 7e-07
 [PIRKW] hormone 4e-07
 [PIRKW] transmembrane protein 2e-06
 [PIRKW] stomach 4e-07
 [PIRKW] actin binding 7e-07
 [PIRKW] ATP 7e-07
 [PIRKW] phosphoprotein 7e-07
 [PIRKW] signal transduction 7e-09
 [PIRKW] heterotrimer 7e-09
 [PIRKW] P-loop 7e-07
 [PIRKW] hydrolase 7e-07
 [PIRKW] transcription regulation 5e-06
 [PIRKW] GTP binding 7e-09

[SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07
 [SUPFAM] SH3 homology 2e-07
 [SUPFAM] SH2 homology 2e-07
 [SUPFAM] protozoan myosin heavy chain IB 7e-07
 [SUPFAM] myosin motor domain homology 7e-07
 [SUPFAM] pleckstrin repeat homology 2e-07
 [SUPFAM] protein-tyrosine kinase src 8e-06
 [SUPFAM] WD repeat homology 3e-12
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07
 [SUPFAM] protein kinase homology 8e-06
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-09
 [SUPFAM] yeast coatamer complex alpha chain 4e-07
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 19
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] Src homology domain 3
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 5.77 %
 [KW] COILED_COIL 2.42 %

SEQ MPTAESEAKVKTKVRFKLLKTHSDLMREKKKKLKKLVLRSEENISPDITIRSNLHYMKETT
 SEGxxxxxxx.....
 COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
 lgotB
 SEQ SDDPDITIRSNLPHIKETTSDDVSAANTNNLKKSTRVTKNKLRLNTQLATENPNGDASVEED
 SEG
 COILS
 lgotB
 SEQ KQGKPNKKVIKTVPOLTTQDLKPETPENKVDSTHQHTKPPQGVHDHOKSEKANEGREET
 SEGxxx
 COILS
 lgotB
 SEQ DLEEDDELMQAYQCHVTEEMAKEIKRKIRKKLEQTYFPSDTLFHDDKLSSEKRRKKKE
 SEGxxxxxxx.....
 COILS
 lgotB
 SEQ VVPFSAETSTLTISGDTVEGEQKKESSVRSVSSDSHQDDEISSMEQSTEDSMQDDTKPK
 SEGxxxxxxx.....
 COILS
 lgotB
 SEQ PKTKKKTKAVADNNDVDGQGVHEITSRDSPPVYPKCLDDDLVLGVYIHRDRLKSDFM
 SEG
 COILS
 lgotB
 SEQ ISHPMKVIHVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTQPYDFKQLKSRLPEW
 SEG
 COILS
 lgotB
 SEQ EEQIVFNENFPYLLRGSDSPKVILFFEILDFLSVDEIKNNSEVQNECGFRKIAWAFK
 SEG
 COILS
 lgotB
 SEQ LLGANGNANINSKRLRLQLYPPPTKPRSPVVEAFEWWSKCPRNHYPTSTLYVTVRGLKVP
 SEG
 COILS
 lgotB
 SEQ DCIKPSYRSMALQEEKGKPVHCHERHHESSVDTEPGLEESKEVIKWKRLPGQACRIPNK
 SEG
 COILS
 lgotB


```

SEQ      HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPIIYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECCCCCEEE

SEQ      SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFHPAVRELVTGTCYDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEETTTEEEETT--TTCEEEETTTCEEEEETTT-TCEEEETT

SEQ      MIRIKVEMREDSAILVRQFDVHKSFINSCLCFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEETTTTTBTTEEEEEECCCCCE-EEEEETTTEEEETTTEEEEEE.....

SEQ      EHSVHHTWINKIKEKTEFKGIPISYLEIHPNGKRLLIHTKDSLRLMDLRILVARKFVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLLFAGSEDIYVWNPETGEQVAMYSDLPFKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAFGQNEPILLYIYDFHVAQQAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FQIDEFVHTESSTKMQLVKQRLETVTEVIRSCAAKVNKNLSFTSPPAVSSQQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQEILHQFGFTQTGIISIERKPCNHQVDTAPTVALYDYTANRSDELTIHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWWYGSIGKGQGYFPANHVASETLYQELPPEIKERSPPLSPEEKTKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      QKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTRMRKNKQAGRKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

Prosites for DKFZphtes3_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005

PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

Pfam for DKFZphtes3_ln3.1

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVWCVaFSPDGGrWFivSGSWDgTCRLWD*
 + GH+N ++++S D ++ I+++ DGT R+W

Query 650 LCGHLNIIYDLSWSKDDHY-ILTSSSDGTARIWK 682

HMM_NAME Src homology domain 3

HMM *pyVIALYDYqAqdpDELSFkEGDIIIIIEdsDD.WWrgRnnnTNGQEGW
 P+V+ALYDY+A+++DEL++ +GDII + ++++ WW+G GQEG+

Query 1054 PTVVALYDYTANRSDELTIHRGDIIRVFFKDNEWWYGSIGK--GQEGY 1100

HMM IPSNYVEPi*

+P+N V+ +
 Query 1101 FPANHVASE 1109

DKFZphtes3_20c21

group: testes derived

DKFZphtes3_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
1 GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGCGCAGGT
51 ACGGGGCGAGA AGTCGCAGGT ACCCAGCTGC TGCCACGTT TCTGGTCCAG
101 AGTCCCAGAAC CCCGAGCACT GGGATGCCTG GCTACTCCGA GCCAAGGCAC
151 TGATGTTTGA ACTGGAACT TCAAAACGTT TAATAAGAGT CTCAGGATG
201 GGTTTGAAGT AGACAAGCTA GAAATTCTTT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTTTGGGCTT TCCTGGAGAG GAGCTTGAAG AGGTGGTTCT
301 GCAGACAGCC ACAGTGATAC TCAGGAAACC AGAGGAATGG ATTTGACTTT
351 TCTGCTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTTGTA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACTTTCC
451 ATGCTGACGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCACGT CCAGATTCAC TGGATTGGGG GATAGAGGAA CGGTCACAGC
551 TAGGGAGAAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAGTCA
651 GCCTCGTGGT GGAATTATTT TTTCTTTTAT GATGGTTCCA AGGTAAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTG TTTACTTTTAT CCTTCCAGA
751 CCCTGCTAGA CCAACAGGAG TTGCTTTGTG GACAGATTGC TGGAGTTGTC
801 CGCTGTGTTT CTGACATTTT TGAATCTCCT CTTACTCTTG TTCGTCTGAG
851 AAAACTGAAG TTTGCCATAA AAGTTGATGG AGATTACCTT TGGGTGCTGG
901 GCTGTGCTGT GGAGCTCCCT GATGTCAGCT GCAAGCGGTT TCTGGATCAG
951 CTAGTTGGAT TCTTTAATTT TTACAATGGA CTGTTTCCC TAGCTTATGA
1001 GAACTGTTCT CAGGAAGAAC TGAGCACGGA GTGGGACACC TTCATCGAGC
1051 AAATTCTGAA AAACACCACT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAACCTAAAGT GGAGCCCTTG TTGTTGCTGA AGGCAGCCCCG
1151 CATTCTGCAG ACCTGCCAGC GCTCGCCTCA CATTCTCGCT GGCTGCATCC
1201 TCTATAAAGG ACTGATTGTC AGCACCCAAC TCCCGCCCTC CCTCACCGCC
1251 AAGGTCCTGC TTCACCGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTCAAGC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAAG
1501 AAAACGCCAC TGGCCATGTG GAATCCATGG CCTGGACCAC CCCAGATCCC
1551 ACATCCCTTG ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACTG CACAACCTG
1651 CCAGGGGTGA GGTCTCTGGC CTCAGCTCCT CCCTGGGGAA GGAACCTAGT
1701 TTTCTCCAAG AAGAATCGA CTTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGAA ATGGCCTCAG GTCATTTTGC CTTCTACAT GTGCCTGTTC
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCCAGCAGC
1851 CTGGAACCCA CGCCTCCTGA GGACACAGCC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTCGAAG
1951 ACCATCCTGG CCATAGCAGC CAAGCCCCCA TTCCAGAGC AGACCCTCTC
2001 CCCAGAAGGA CCGCAGGCC CTTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGGAAAC AAGCTTCCCA CGGGGGAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGGCCAC GCAGCCCTTG GTCTGGAATG CAGTTCAGGC
2151 TCAGCAAACCT GTCAGGGTGC TGGCCCTCT GCAGATGGAA TCAGCTCCAG
2201 GCTGACACCA GCAGAGTCTT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACTGCGT CAAAGGGCTG ATGCTGTCCC TGCTGGCTGA GGAGCCGCTG
2301 CTGGGAGACA GCGCAGCCAT AGAGGAAGTG TACCACAGCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAAGA GACGCTGCCC AGGGATGAGG
2401 CAGCCTCCAC GAGCAGCACC TACAACCTCA CATATTACGA CCGCATTCAG
2451 AGCTTGCTGA TGGCAAACTT GCCGCAGGTG GCCACCCCGC ATGATCGCCG
2501 CTTCTCCAG GCCGTACGCC TGATGCATAG CGAATTGCCC CAGTGCCTCG
2551 CGCTTTATGA AATGACTGTC AGAAATGCCT CCACGGCTGT GTACGCTGT
2601 TGCAACCCCA TCCAGGAGAC ATATTTCCAG CAGCTGGCAC CTGCAGCAG
2651 GAGTCCGGC TTCCCAAACC CTCAGGATGG CGCCTTCAGC CTCTCCGGCA
2701 AAGCAAAGCA GAAGCTGCTG AAGCACGGGG TGAACCTGCT CTGAACTGCA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAGGAAG CTCTGCCTTT
2801 TATAATTGAA AAGGCCCTCT TATTTTATTT TTCTTGAAAA CATTCCCTTT
```



```

2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG GTTAGTTCTT CAGGTCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACTGTATT TAATCGGTAG TGTGGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTCATA TCTCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCACAGT GGCATTTTTT
3101 CTAGAACCCTG ATCGTCCACT AGCCCAGAGT GTGTGTGTTT AACCCACACA
3151 CCAGGTGGTG GTAGGCGGTG TGAAGTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGCAGGT AGGTTTCTCC AGTGCCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTTG AATTCTGTG GAACTCTACT CAAATCTAGG GCGCTCTTCT
3351 TTGACCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CTTTATAAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCTGC GCTGCCTTTG ACTACAAGTG GGCCGTGCAG
3501 CTGGTTCTCT TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGTCTG
3551 GGGGTAAAAG CACTGTGCTT TTCAGTGGTG GCTGCGTGAA AGGGAGCGAC
3601 ACTCAGCTGT GTGTCTCTGG GCTTGTGTGG TACTTAGAAC CTCAGTTCTA
3651 TTACGTTATA GTCAGACATT TTTTGGACAG TATGAGACAG ACTGCAGGAT
3701 GAAAAATATT GTCAAAATCT TAACTGAATG TTTACTGGAA GACTTGAGA
3751 TTCCATTGTA GAGTTGTATT GTTAATAATT TCATGTCAGT GAACTGATAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HS1048E9 from database EMBLNEW:
 Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2
 Contains pseudogene similar to ribosomal protein S3A and part of a gene
 similar to C.elegans protein CE02118, ESTs, STS, GSS.
 Score = 6540, P = 0.0e+00, identities = 1308/1308
 -14 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708
 Category: putative protein
 Classification: no clue

```

1 MATSTSTEAK SASWNYFFL YDGSVKKEEG DPTRAGICYF YPSQTLDDQQ
51 ELLCGQIAGV VRCVSDISDS PPTLVRLRLK KFAIKVDGDY LWVLGCAREL
101 PDVSKRFLD QLVGFNFYFN GPVSLAYENC SQEELSTEDW TFIEQILKNT
151 SDLHKIENSL WNLDQTKVEP LLLLKAARIL QTCQRSFIL AGCILYKGLI
201 VSTQLPPSLT AKVLLHRTAP QEQLPTGGD APQEHGAALP PNVQIIPVFV
251 TKEEAILLHE FPVEQMTRSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWTTTPD PTSPDEACPD GRKENGCLSG HDLESIRPAG LHNSARGEVL
351 GLSSSLGKEL VFLQEELDL EIHIEPAQEV EMASGHFAFL HVPVPDGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPPSA PEMLTQHGAQ EQVEDHDPGHS
451 SQAPIPRADP LPRRTTRPL LPRLDPGQRG NKLTPTGEQGL DEDVDGVCEG
501 HAAPGLECSS GSANCOGAGP SADGISSRLT PAESCMGLVR MNLYTHCVKG
551 LMLSLAEEP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTYYDRI QSLLMANLPQ VATPHDRRL QAVSLMHSEF AQLPALYEMT
651 VRNASTAVYA CCNPIQETYP QQLAPARSS GFNPQDGAFL SLGKAKQKQL
701 LKHGVNLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_20c21, frame 3

Report for DKFZphtes3_20c21.3

692

DKFZphtes3_20k2

group: signal transduction

DKFZphtes3_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
101 GTGATGGAGA GTCTCTGCCG TGCCATCTGG GATGCAAACC GTCCCTGTGT
151 CCCCCACGTC CAGGCCGTAG ATGCTCCCCG CCGGTCAGTC ACTTAGTCGT
201 CAGATCGCCC GTCTTGGTAT CACAGTGCTT CTGTTCAAGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGGG GACCCACTCC AAAAGGACAC CTGCCAGAC CCCCTGGATG
351 GAGACCCATA CTCCAGGCCA CCTCCAGCCA AGCCCCAGCT CTCCACGGCC
401 AAGAGCCGCA CCCGGCTCTT TGGGAAGGGT GACTCGGAGG AGGCTTTCCC
451 GGTGGATTGC CCTCACGAGG AAGGTGAGCT GGACTCTTGC CCGACCATCA
501 CAGTCAGCCC TGTATCACC ATCCAGAGGC CAGGAGACGG CCCACCCGCT
551 GCCAGGCTGC TGTCCCAGGA CTCTGTCTGC GCCAGCACCG AGAAGACCCT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCGTTGCT CAGAATAACT
651 GCCAGGATCT GGAGAGCCTG CTGCTCTTCC TGCAGAGAG CAAGAAGCAC
701 CTCACAGACA ACGAGTTCAA AGACCCTGAG ACAGGGAAGA CCTGTCTGCT
751 GAAAGCCATG CTCACCTGCG ATGACGGACA GAACACCACC ATCCCCCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCACCGCC
851 AGCTACACGG ACAGCTACTA CAAGGGCCAG ACAGCACTGC ACATCGCCAT
901 CGAGAGACGC AACATGGCCC TGGTGACCTT CCTGGTGGAG AACGGAGCAG
951 ACGTCCAGCT TCGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAGGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTTG TCCTTGCCCG GTGCACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGTCTGC GAACTCTTGG CAGACGGCCG
1101 ACATCAGCGC CAGGGACTCG GTGGGCAACA CCGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACACG AAGTTTGTGA CAGGATGTA
1201 CAATGAGATT CTGATCCTGG GGGCCAAACT GCACCCGACG CTGAAGCTGG
1251 AGGAGCTCAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGTCTT GGCCTATATT CTCCAGCGGG AGATCCAGGA
1351 GCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTCGCTGTAC GACCTGTCTT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCCT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTGCGCTCCC TTAAAGATGG AAAAAATTGG AGACTATTTC CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTT CCGAGGGATT
1751 CAGTATTTCC TGCAGAGGCG GCCGTCGATG AAGACCCTGT TTGTGGACAG
1801 CTACAGTGAG ATGCTTTTCT TTCTGCAGTC ACTGTTTCATG CTGGCCACCG
1851 TGGTGCTGTA CTTCAAGCAC CTCAAGGAGT ATGTGGCTTC CATGGTATTC
1901 TCCCTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTTGTG TACATCGTCT TCTTGTTCGG GTTTTCCACA
2051 GCGGTGTTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTCG CACAGGTGGC GGGGGCCTGC CTGCAGGCCC CCCGATAGCT
2151 CCTACAACAG CCTGTACTCC ACCTGCCTGG AGCTGTTCAA GTTCACCATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCTG CTGCTGGCCT ATGTAATTCT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGGAG
2351 AGCAAGAACA TCTGGAAGCT GCAGAGAGCC ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCCGCTCA GGCAAGCTGC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACTG GACCACCTGG AACACCAACG TGGGCATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGGCGT CAAGCGCACC CTGAGCTTCT
2601 CCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCTCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```



```

2701 CGAGGAAGTT TATCTGCGAC AGTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCCTTAG GAGACCCCGT TGCCACGGGG
2851 GGTGCTGAG GGAACACCAG TGCTCTGTCA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCAGC ATGTTCCCAA ATCTGTGCTG GACAAGCTGT GGGGAAGCGT
2951 CTTGGAAGCA TGGGGAGTGA TGTACATCCA ACCGTCACGT TCCCCAAGTG
3001 AATCTCCTAA CAGACTTTCA GGTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTC TTTGATTTTA
3101 TTCTTTTTTT TGAGACAGAA TTTCACCTT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCTGCCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGCTGGCTA ATTTTTTGA TTTTTTAAAT AGATATGGGG TTTGCCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATCTTTT TTTTTTTTCT GTGAGACAGA GTTTCACCTC
3451 TGTGCCCCAG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCCTCCCGG GTTCAAGCGA TTCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGGT GAGCACTACC ACGCCCGCT AATTTTTGTA TTTTAAATAG
3601 AGACGGGGTT TCACCATGTT GGCCAGGCTG GTCTCGAAGT CTTGACCTCA
3651 GGTGATCTGC CCGCCTTGGC CTCCCAAAGT GCTGGGATTA CAGGTGTGAG
3701 CCGCTGCGCT CGGCTTCTT TGATTTTATA TTATTAGGAG CAAAAGTAAA
3751 TGAAGCCAGC GAAACACCT TTGGGAACAA ACTCTTCCTT TGATGGAAAA
3801 TGCAGAGGCC CTTCTCTCT GTGCCGTGCT TGCTCCTCTT ACCTGCCCGG
3851 GTGTTTTGGG GGTGTTGGTG TTTCTCCCT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTCTGGCA GAATCAAGCT GTTGACAGCAG TGCCTTCTTC
3951 ATCCTTCTT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGATGT
4051 TTGAAGATC TTCCATTCT AGGAACCCCA GTCCTGCTT TCCGCAATGG
4101 CACATGCTTC CACTCCATCC ATACTGGCAT CCTCAAATAA ACAGATATGT
4151 ATACATATAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

99288727:
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:
A non-pungent triprenyl phenol of fungal origin, scutigeral, stimulates rat dorsal root ganglion neurons via interaction at vanilloid receptors.

Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839
Category: strong similarity to known protein
Classification: Cell signaling/communication

```

1 MKKWSSTDLG AAADPLQKDT CPDPLOGDPN SRPPPAKPOL STAKSRTRLF
51 GKGDSEEAFF VDCPHEEGEL DSCPTITVSP VITIQRPGDG PTGARLLSQD
101 SVAASTEKL RLYDRRSIFE AVAQNQCQL ESLLLFLQKS KHLTDNEFK
151 DPETGKTCIL KAMNLHDGQ NTIPLLEI ARQDLSLKEI VNASYTDSYY
201 KGQALHIAI ERRNMALVTL LVENGADVQA AAHGDFEKKT KGRPGFYFGE
251 LPLSLAACTN QLGIVKFLQ NSWQTADISA RDSVGNVTLH ALVEVADNTA
301 DNTKFVTSY NEILILGAKL HPTLKEELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECRHLRKE TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 MAAYYRPVDG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFFLQS LFMLATVVLY FSHLKEYVAS MVFSLALGWT
551 NMLYYTRGFQ QMGIIYAVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGKNDLPSSE STSHRWGPA CRPPDSSYNS LYSTCLELFK FTIGMGDLEF
651 TENYDFKAVF IILLAYVIL TYILLNMLI ALMGETVNKI AQESKNIWKL
701 QRAITILDTE KSFLKCMRKA FRSGKLLQVG YTPDGKDDYR WCFRVDEVNW
751 TTWNTNVGII NEDPGNCEGV KRTLSFSLRS SRVSGRHWKN FALVPLLRFA
801 SARDRQSAQP EEVYLRQFSG SLKPDAEVF KSPAASGEK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20k2, frame 2

TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBLNEW:AB015231_1 product: "stretch-inhibitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhibitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds.
Length = 838

HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query:      1 MKKWSSTDLGAAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
             M++ +S D   + P Q+++C DP D DPN +PPP KP + T +SRTLFGKGDSEEA P
Sbjct:      1 MEQRASLDSESESPQENSCLDPPDRDPNCKPPPVKPHIFTRSRTRLFGKGDSEEAFF 60

Query:     61 VDCPHEEGELDSCTITVSPVITQRPDGPPTGARLLSQDSVAASTEKTLRLYDRRSIFE 120
             +DCP+EEG L SCP ITVS V+TIQRPDGP R SQDSV+A EK RLYDRRSIF+
Sbjct:     61 LDCPYEEGGLASCIITVSSVLTIQRPDGPASVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query:    121 AVAQNNCQDLESLLLFLQKSKKHLTDNEFKDPETGKTCLLKAMNLNLDGQNTTIPLLEI 180
             AVAQ+NCQ+LESLL FLQ+SKK LTD+EFKDPETGKTCLLKAMNLN+GQN TI LLL++
Sbjct:    120 AVAQNCQDLESLLPFLQSKKRLTDSEFKDPETGKTCLLKAMNLNHNQNDTIALLLDV 179

Query:    181 ARQTDLSKELVNASYTDSYYKGQTAHIAIERRNMALVTLLEVAGADVQAAAHGDFFKKT 240
             AR+TDSLK+ VNASYTDSYYKGQTAHIAIERRNM LVTLLVAGADVQAAA+GDFFKKT
Sbjct:    180 ARKTDLSLKQFVNASYTDSYYKGQTAHIAIERRNMTLVTLLEVAGADVQAAANGDFFKKT 239

Query:    241 KGRPGFYFGELPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNTVLHALVEADNTA 300
             KGRPGFYFGELPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNTVLHALVEADNT
Sbjct:    240 KGRPGFYFGELPLSLAACTNQLAIVKFLQNSWQPADISARDSVGNTVLHALVEADNTV 299

Query:    301 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKGMTPLALAAGTGKIGVLAYILQREIQE 360
             DNTKFVTSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct:    300 DNTKFVTSMYNEILILGAKLHPTLKLEETNRKGLTPLAALAASSGKIGVLAYILQREIHE 359

Query:    361 PECHRLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEIAYSSSETPNRHDMMLVEPLN 420
             PECHRLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEIAYSSSETPNRHDMMLVEPLN
Sbjct:    360 PECHRLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEIAYSSSETPNRHDMMLVEPLN 419

Query:    421 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVDGLPPFKMEK-IGDYFRVTGEI 479
             RLLQDKWDRFVKRIFYFNFLVYCLYMIIFT AAYYRPV+GLPP+K++ +GDYFRVTGEI
Sbjct:    420 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTAAAYYRPVEGLPPYKLKNTVGDYFRVTGEI 479

Query:    480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSHLKEYVA 539
             LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+QSLFML +VVLYFS KEYVA
Sbjct:    480 LSVSGGVYFFFRGIQYFLQRRPSLKSFLVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query:    540 SMVFSLALGWTNMLYYTRGFQQMGIYAVMIEKMILRDLCREMFVYIVFLFGFSTAVVTLI 599
             SMVFSLA+GWTNMLYYTRGFQQMGIYAVMIEKMILRDLCREMFVY+VFLFGFSTAVVTLI
Sbjct:    540 SMVFSLAMGWTNMLYYTRGFQQMGIYAVMIEKMILRDLCREMFVYVFLFGFSTAVVTLI 599

Query:    600 EDGKNDSLPESTSHRWGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 659
             EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV
Sbjct:    600 EDGKNNSLPMESTPHKCRGSACKP-GNSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 658

Query:    660 FIILLLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDEKSFLKCMRK 719
             FIILLLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDEKSFLKCMRK
Sbjct:    659 FIILLLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDEKSFLKCMRK 718

Query:    720 AFRSGKLLQVGYPDGGDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSRLR 779
             AFRSGKLLQVG+TPDGGDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSRLR
Sbjct:    719 AFRSGKLLQVGFTPDGGDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSRLR 778

Query:    780 SSRVSGRHWKNFALVPLLREASARDRQSAQPEEVYLRQFSGSLKPDAEVFKSPAASGEK 839
             S RVSGR+WKNFALVPLL+AS RDR + Q EEV L+ ++GSLKPDAEVFK GEK
Sbjct:    779 SGRVSGRNWKNFALVPLLRASTRDRHATQEEVQLKHYTGSLKPDAEVFKDSMPVGEK 838
```

Pedant information for DKFZphtes3_20k2, frame 2

Report for DKFZphtes3_20k2.2

[LENGTH] 839
 [MW] 94950.75
 [PI] 6.90
 [HOMOL] TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus
 vanilloid receptor subtype 1 mRNA, complete cds. 0.0
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05
 [PIRKW] alternative splicing 3e-06
 [PIRKW] peripheral membrane protein 3e-06
 [SUPFAM] ankyrin repeat homology 3e-06
 [SUPFAM] unassigned ankyrin repeat proteins 3e-06
 [PFAM] Ank repeat
 [KW] TRANSMEMBRANE 4

```

SEQ  MKKWSSTDLGAAADPLQKDTCPDPLDGDPSRPPPAKPQLSTAKSRTLFGKGDSEEAFF
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  VDCPHEEGELDSCPTITVSPVITIQRPGDPTGARLLSQDSVAASTEKLRLYDRRSIFE
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  AVAQNNCQDLESLLLFLQSKKHLTDNEFKDPETGKTCLLKAMLNLDHGQNTTIPLLLEI
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  ARQTDLSKELVNASYTDSYKQGTALHIAIERRNMALVTLLEVENGADVQAAAHGDFKKT
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  KGRPGFYFGLPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNTVLHALVEVADNTA
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQREIQE
PRD  chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  PECRHLRKFTEWAGVPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPPVDGLPPFKMEKIGDYFRVTGEIL
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  SVLGGVYFFFRGIQYFLQRRPSMKTFLVDSYSEMLFFLQSLFMLATVVLYFSHLKEYVAS
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  MVFSLALGWTNMLYYTRGFQOMGIYAVMIEKMILRDLCRFMFVYIVFLFGFSTAVVTLIE
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  DGNDSLPSSESTSHRWGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAVF
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MM

SEQ  IILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSLKCMRKA
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  FRSGKLLQVGYPDGGDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLRS
PRD  hhceeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  SRVSGRWKFNALVPLLEASARDRQSAQPEEVYLRQFSGSLKPEDAEVFKSPAASGEK
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

```

(No Prosite data available for DKFZphtes3_20k2.2)

Pfam for DKFZphtes3_20k2.2

HMM_NAME	Ank repeat		
HMM		*GyTPLHIAARYNNvEMVr1LLQHGDIN*	
		G+T+LHIA +++N+ +V LL+++GAD+	
Query	202	GQTALHIAIERRNMALVTLLVENGADVQ	229

DKFZphtes3_2013

group: transmembrane protein

DKFZphtes3_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp

Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCTGCGTT GTTGTCACT GGAGAGCAGG GAGTGGGGCC
51 AGCCAGCAGA AACAGTGGGC TGTACAACAT CACCTTCAAA TATGACAATT
101 GTACCACTTA CTTGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCCAG
151 AATATACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCCTGAAA GGATTTCGGG
251 TAATACTGGA GGAGCTGAAG TCGGAGGGAA GACAGTGCCA ACAACTGATT
301 CTAAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCTGAATA TGAATTTGA AACGGATTAT TTCGTAAAGG
401 TTGTCCCTTT TCCTTCCATT AAAAACGAAA GCAATTACCA CCCTTTCTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGGACA ATCTAGCTTG
501 TAAACCCCTT TGGAGCCTC GGAACCTGAA CATCAGCCAG CATGGCTCGG
551 ACATGCAGGT GTCCTTCGAC CACGCACCGC ACAACTTCGG CTTCCGTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCAGCAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
701 TTTCTCCAGG GGATTATATA ATTGAGCTGG TGGATGACAC TAACACAACA
751 AGAAAAGTGA TGCATTATGC CTTAAAGCCA GTGCACTCCC CGTGGGCCGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGGTAGTC ATATCGGCAT
851 TCGCGACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAAATATA
901 TATTACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT ACACTGCAGC
951 ACTCCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGGTCTTT CTCTGCTATT
1001 CCAAGTAAAG TGGCCAGAAT CACATGAATG TCGTCCAGTG TTTCCGCTAC
1051 TTCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTCAGCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTCATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGT CCAAAGGTAT GAAGTACTTT
1201 GTGGACAAGA AGAACTACAA ACACAAAGGA GGTGGCCGAG GCTCGGGGAA
1251 AGGAGAGCTC TTCCTGGTGG CGGTGTCAGC CATTGCCGAA AAGCTCCGCC
1301 AGGCCAAGCA GAGTTCGTCC GCGGCGCTCA GCAAGTTTAT CGCCGTCTAC
1351 TTTGATTATT CCTGCGAGGG AGACGTCCCC GGTATCCTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCCTCA GCTCTGTTC CACCTGCACT
1451 CCCGAGACCA CGGCCTCCAG GAGCCGGGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAACT ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACGTCGCCAT
1551 TTGCAACATG CACCAAGTTA TTGACGAGGA GCCCGACTGG TTCGAAAAGC
1601 AGTTCGTTC CTTCCATCCT CCTCCACTGC GCTACGGGGA GCCAGTCTTG
1651 GAGAAATTG ATTCTGGGCTT GGTTTAAAT GATGTCATGT GCAAACCAAG
1701 GCCTGAGAGT GACTTCTGCC TAAAGGTAGA GGCGGTGTT CTTGGGGCAA
1751 CCGGACCAGC CGACTCCAG CACGAGAGTC AGCATGGGGG CCTGGACCAA
1801 GACGGGGAGG CCCGGCCTGC CCTTGACGGT AGCGCCGCCG TGCAACCCCT
1851 GCTGCACACG GTGAAAGCCG GCAGCCCTC GGACATGCCG CGGGACTCAG
1901 GCATCTATGA CTCGTCTGTG CCCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCTCTCTCT TCAGGCCTGG GTGAGGAGGA ACCTCCTGCC CTTCTTCCA
2051 AGTCCTCTC TTCTGGGTCA TGCAAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGCGGT CGCCCTTTG TAACAAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTAGCTGCG TGCTCCCTC TGATTCCCCA GCTCATCTCC
2201 CTGGTTGCAT GGCCCACTTG GAGCTGAGGT CTCATACAAG GATATTTTGA
2251 GTGAAATGCT GGCCAGTACT TGTCTCCCT TGCCCCAACC CTTTACCGGA
2301 TATCTTGACA AACTCTCCAA TTTTCTAAAA TGATATGGAG CTCTGAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2401 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595
 Category: similarity to known protein
 Classification: unclassified

```

1 MESQPFLLNMK FETDYFVKVV PFPSIKNESN YHPFFFRTRA CDLLLPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKHGEPFKR
101 KTKQEQTE MTSCLLQNVS PGDYIIELVDTNTTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFATLFTVMCRKK QENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNHMNVV QCFAYFLQDF CGCEVALDLW
251 EDFSLCREGQ REWVIQKIHE SQFIIVVCSK GMKYFVDKKN YKHKGGRGS
301 KGELFLVAV SAIAEKLRA KQSSAALSK FIAVYFDYSC EGDVPGILD
351 STKYRLMDNL PQLCSHLHSR DHGLQEPGQH TRQSSRRNYF RSKSGRSLYV
401 AICNMHQFID EEPDWFEKQF VPFHPPPLRY REPVLEKFD GLVLNDVMCK
451 PGPESEDFCLK VEA AVLGATG PADSQHEQHG GGLDQDGEAR PALDGSAAQ
501 PLLHTVKAGS PSDMPRDSGI YDSSVPSEL SLPLMEGLST DQTETSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGSKADL GCRSYTDELH AVAPL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2013, frame 1

TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
 mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus
 interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =
 1.1e-13

>TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
 mRNA, complete cds.
 Length = 866

HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14
 Identities = 85/284 (29%), Positives = 131/284 (46%)

```

Query: 213 KVFLCYSSKDGQNHMNVVQCFAYFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
      KV++ YS+ D +++VV FA FL CG EVALDL E++ G WV QK +
Sbjct: 379 KWIYYSA-DHPLYVDVVLKFAQFLLTACGTEVALDLLEEQAISEAGVMTWVGRQKQEMV 437

Query: 269 HESQFIIVVCSKGMKY----FVDKKNXXXXXXXXXXXXXELFLVAVSAIAEXXXXXXXX 324
      + IIV+CS+G + + + +LF A++ I
Sbjct: 438 ESNSKIIIVLCSRGTRAKWQALLGRGAPVRLRCDHGKPVGDLFATAMNMILPDFKRPACFG 497

Query: 325 XXXXXXFIIVYF-DYSCGVDVPGILDSTKYRLMDNLPQLCSHLHSRDHGLQEPGQHTRO 383
      ++ YF + SC+GDVP + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGDVDFLGAAPRYPLMDRFEEV--YFRIQDLEMFQGRMHRV 550

Query: 384 G--SRRNYFRSKSGRSLYVAICNMHQFIDEEPWFQKQFV----PFHPPPLR---YREPV 434
      G S NY RS GR L A+ + PDWFE + + P L + EP+
Sbjct: 551 GELSGDNLYRSPGGRQLRAALDRFDWQVRCDFECENLYSADDQDAPSLDEEVFEEPL 610

Query: 435 LEKFDGLVLNDVMCKPGPESEDFCLKVEAVLGATGPADSQHEQHGGLDQDGEARP 491
      L +G+V + + P S CL ++ V G GA H L G+ P
Sbjct: 611 LPP-GTGIVKRAPLVRE-PGSACLAIDPLV-GEEGGAIVAKLEPH--LQPRGQAP 662

```

Pedant information for DKFZphtes3_2013, frame 1

Report for DKFZphtes3_2013.1

[LENGTH] 595
[MW] 66847.05
[pI] 6.27
[HOMOL] TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus interleukin
17 receptor mRNA, complete cds. 2e-14
[BLOCKS] BL00740A MAM domain proteins
[BLOCKS] BL01224B N-acetyl-gamma-glutamyl-phosphate reductase proteins
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 13.61 %

SEQ MESQPFNLNMFETDYFVKVVPFSPKSNESNYHPFFRTRACDLLQPDNLACKPFWKPRN
SEG
PRD ccc
MEM

SEQ LNIHQHSDMQVSFDHAPHNFGFRFFYLHYKLKHEGPFKRKTKQEQTTEMTSCLLQNV
SEG
PRD eeeeecc
MEM

SEQ PGDYIIELVDDTNTTRKVMHYALKPVHSPWAGPIRAVAITVPLVVISAFATLFTVMCRKK
SEG
PRD ccc
MEMMMMMMMMMMMMMMMMM.....

SEQ QQENIYSHLDEESSESSTYTAALPRERLRPRKVFCLCYSSKDGQNHMVVQCFAFLQDF
SEG
PRD hhh
MEM

SEQ CGCEVALDLWEDFSLCREGQREWVIQKIHESQFIIVCSKGMKYFVDKKNYKHKGGRGS
SEG
PRD cchhh
MEM

SEQ GKGEFLVAVSAIAEKLQAKQSSAALSKEFIYVFDYSCEGDVPGILDSTKYRLMDNL
SEG
PRD ccc
MEM

SEQ PQLCSHLHSDHGLQEPGQHTRQGSRRNYFRSKSGRSLYVAICNMHQFIDEEDWFEKQF
SEG
PRD cchhh
MEM

SEQ VPFHPPFLRYREPVLKFDGLVNDVMCKPGPESDFCLKVEAAVLGATGPADSQHEQ
SEG
PRD ecc
MEM

SEQ GGLDQDGEARPALDGSAAQLLHTVKAGSPSDMPROSGIYDSSVPSELSPLEGLST
SEG
PRD ccc
MEM

SEQ DQTETSSLTESVSSSSGLGEEPPALPSKLLSSGSKADLGCRSYTDELHAVA
SEG
PRD hhh
MEM

(No Prosite data available for DKFZphtes3_2013.1)

(No Pfam data available for DKFZphtes3_2013.1)

DKFZphtes3_20m18

group: nucleic acid management

DKFZphtes3_20m18 encodes a novel 132 amino acid protein with similarity to the *S. cerevisiae* mitochondrial carrier protein RIM2.

The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer. The RIM2/MRS12 gene encodes a predicted protein of 377 amino acids that is essential for mitochondrial DNA metabolism and proper cell growth. Inactivation of this gene causes the total loss of mitochondrial DNA and, compared to wild-type rho⁰ controls, a slow-growth phenotype on media containing glucose. The novel protein seems to be the human orthologue of this protein.

The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

similarity to carrier protein RIM2

Sequenced by MediGenomix

Locus: unknown

Insert length: 3572 bp

Poly A stretch at pos. 3530, polyadenylation signal at pos. 3510

```
1  GCGCGGGGA GGGCTGTGCC GGTGCTTTC TGCAGCCGCA TCTCGGCCAG
51  CTCTCCTCGC CGTCCCCGGG GCGCTGTGCG TCTCCAGTCC GGGACCGAAG
101 CCGCCTGCCG TAGCGGGCGG CCAGATCCGC GTCCCGCCTC AGCGGCCGGA
151 GGACATGCGG GAGAGAGAAT GAGCCAGAGG GACACGCTGG TGCATCTGTT
201 TGCCGGAGGA TGTGGTGGA CAGTGGGAGC TATTCTGACA TGTCCTACTGG
251 AAGTTGTAAA AACACGACTG CAGTCATCTT CTGTGACGCT TTATATTTCT
301 GAAGTTCAGC TGAACACCAT GGCTGGAGCC AGTGCAACC GAGTAGTGTC
351 TCCCGGACCT CTTCAATTGC TAAAGGTGAT CTTGGAAAA GAAGGGCCTC
401 GTTCCTTGTT TAGAGGACTA GGCCCCAATT TAGTGGGGGT AGCCCCCTTC
451 AGAGCAATAT ACTTTGCTGC TTATTCAAAC TGCAAGGAAA AGTTGAATGA
501 TGTATTTGAT CCTGATTCTA CCCAAGTACA TATGATTTC GCTGCAATGG
551 CAGGTATGAA TGTATAATAT TAAAAAATAA AAAAAGTTTC TGAAACCTAG
601 AGGCTTAATA TTGAATTATA AGTTTGTAGT GAAAAGTTGA TGATTAATGT
651 GCTTTTCATT GATTAGATGA TTTTACGTT TATCGATATA AACCAAATTA
701 GGTATATGTA AAATCTGTCA TCAGTTGACA TTTTGTAGT CAGGAGTTTA
751 CATGCTAGGG TACAAGTAAT ATATTTATAT TGCCTTGTGT AGTCCACTGA
801 ATGTTTAGTG ATCATTGTTA ACAGTTTAA GAATCCAACC ATAATTACAC
851 TATAAATAAG TTATGGAGCT GTAATTTACT CTTCTCTCCT CAATTTCTGT
901 TAGTGCCCTT TCCCTTTTGG CTGCATGTTT TGGCTTCTGT CTGAAATGTG
951 TCGGCAATTC TTGGTAAAGT ATTCATTTTG TCCTGTGCTC AAATGCTGAA
1001 ATTTTGTGTA GTGATGTATT ATTATTGACA ATTCAGTTAC TATGTGTATT
1051 TTTTAAATTT GTTTATTATT CTACATAATT CACACTAGAC AGCACCTGAA
1101 ATTTAGACAC TGGCTATGTG TACATGCTTA CTATAGAAAT GTTTCAGGA
1151 ACTCTCTGTT TCTGTATCA CTGATAAGTA TATATGATTC TGAATTAATA
1201 TAACTAGTTT TAGGTCTTTA CCCTGCCATA AAGATAAACA GTTGGTTTGA
1251 CCAATCTGGT TCTGGAATCA TTTGCTGCTA TGCATGTTAG ACAAGGCCAC
1301 GAACCTTGAT TTTCCATTGA AAATCTCTCC TAATATCTGA GATTTATTGT
1351 ATATTTACTC ATATCTCACA TTTTCAAATT ATGCTGTAAC TTTATAAACT
1401 GTAGCTGCTT TCATCAGCTA TTGATCAATA AATTGAATGT CAATTATGTG
1451 CTTAATAATG AGTGCCTTAA ACTGTTAAAC ACTTTTGGTT TAGAAATAAA
1501 GTGAATCAAT TTGACCTATA TACTTCATGA AGTAAGTAAG TTTGAAATAC
1551 AAATTTCTGA AAGGTCAATA GCCCTTATCG TATTACAAAT TGTTTTTAAG
1601 GCTTTTGTGA TTTATTAATT GTCAGTTGAT TCACTGAAGC TTTAAACTG
1651 GAAGGGACAA TCCAAAGGTC AAAAGAGTGA AATACAATCA TTTACCAATA
1701 AGGAAACCTT GGGCAAATTA TGTAATTTAT GTGAACCTCT CTTAGCTTAC
1751 CCATGGAATG AGTCAAGTGG TCTACATAGA TTTGGATTTT GAGAATTAGT
1801 TCTTTTCATT AGTGTTATAG AGATTATCTT GTTACAACCTA GAATTATTTT
1851 TAATGTAATT TTTACAGATG TTGAATATTA GTAGATAGGA TTTTCCCCT
1901 ACGAATTTGG ATGTAAGGTA AAGGTTGGTG GCCAGTGACA AACCTTATAA
1951 CCACCTTTATC AGGTTCTTTA AAAATATATT TGTGAATTAC CAGTGATTAT
2001 GTTTTGGGCT TATAACCTCA GATAATTATA AAGAAATGTT AATCTTATTT
2051 GAAAGAATTG GAATCTAGAA AGTTAGATGA GCAGTCATTT TATATTGATA
2101 TTTGTTATAT CAGTATAGCA AATGCAGAGG TTCAGAATAT CTTTATTTCC
2151 ACTGGAACAT CTTATTTTCT TAGAGTATCT CATCAGAAAT TATTACTGTA
2201 TTTGTATCAC ATTGCAAGA ATTTCACTAG AATTGTCAGT TTGCACTTTT
2251 TTCTCAAAATG TGTACAAATG TTAACATATA GTTCATTTTT ATCTGTACAT
2301 TGATGCCATT TCCCAACTTG AATTCCTCAA GTTTTGGTAA ACTTACAATC
2351 TCATACTTGT TCAGAGGTTA TTGCACTGTA CACTTACTGT GTAGAAAATA
2401 CTGTTTGAAT TTGTTTGCAG TTACATTGTT CTGAGAACTG TGCTCTCAGA
2451 GCTTCTGTGC ACTATTCATG AGCATTAAAC CTTAGCCTTG CAGTTTTATA
```



```

2501 CATAACTATA TGGTTAGTAA AACTGAATGG TCCAATGCAG ACTCATTAAA
2551 GTAGGCTTTT GCCCCTTTG TTCTTGAAAT AATCTAGACC AGATTACTCG
2601 GGGTTTTTTT TAGGATTATT TTTATAGGTC TAAATATGAA TGATTTGGGG
2651 GTATGAAGTA CTAAAGATA GTTCTGTGAA AAATCATTTT CAGCTGTCTA
2701 TTCAAGGGAA AAAATGCTAA CCTTGTCAC TACTACACA AAACCACACT
2751 AAAATAAACC ATTAATGATA CTGCCTGCAA GATTTTAACA CACCAGATAG
2801 CACACACATT AAGGATTTAT AAGGCACTGT ACGTAATTTT TATCCAAGT
2851 GACCTCTCAA TTCATTTTCA TTTTGCAATT TATCCATATG AACTCATGTT
2901 TAATTTAGAT AATAAAAATT TATTTTATTA AAAGGACAGT TTATTTAAAG
2951 TGGGTCCTTT TATTTGTTGT AGTGCATACT ATAAGAATTT GTAAGCCTCT
3001 AAAGTTGAGC TATAAATTTT CATGCATTAA AAATTTGTTT CAGTTGTGAG
3051 GATATTTAAT CAGATTAAAT AATGTTGACT CTTAATATTT TGCTGCCTT
3101 TTTTTCCTCC TACACATGAC CTTTGACAGA CTAAGTATAT CTCAGCTATT
3151 GAGGGTATCT GTTTTGTGTC CTGTATATTT TGTTTAAATT AACTTGATA
3201 TTCCTTTGTA TACACCTAGG CACAGATGTA TGCAAAAAAA ATTTGTTAAA
3251 TTAATCTCTT CTTTATACTA ATTCTCAATT TTTAAAAGAT TTTATCTGGC
3301 ATGTATATAC TTTTATATAG AACATTATAA ATGTAAAGGA AATGAATTCT
3351 AATTTTAATT GGATTATGTA TTCATACAGT TATCTCAAT TTTTAAATA
3401 CTAATAATGT AATCATTGAA TGTTTCCTAC ATACGTAGTG GGTTTTATTT
3451 GCTCACAGCA TACAGTTATT TTTCAATTTA TGTTTTCTTA TTAGACTTAA
3501 ATTTCAATTAT AATAAAGGCT TTTACTCATT AAATACAAAA AAAAAAATAA
3551 AAAAAAATAA AAAAAAATAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

95198680:

Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria.

Peptide information for frame 1

ORF from 169 bp to 564 bp; peptide length: 132
 Category: similarity to known protein
 Classification: Intracellular transport and traffic
 Prosite motifs: LEUCINE_ZIPPER (27-49)
 MITOCH_CARRIER (26-36)

```

1 MSQRDTLVHL FAGGCGGTVG AILTCPLEVV KTRLQSSSVT LYISEVQLNT
51 MAGASVNRVV SPGPLHCLKV ILEKEGPRSL FRGLGPNLVG VAPSRAIYFA
101 AYSNCKEKLN DVFDPDSTQV HMISAAMAGM NV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZpHtes3_20ml8, frame 1

PIR:S44092 probable carrier protein c2 - *Caenorhabditis elegans*, N = 2,
 Score = 147, P = 1.5e-19

PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*), N = 1, Score = 230, P = 6.2e-19

>PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*)
 Length = 377

HSPs:

Score = 230 (34.5 bits), Expect = 6.2e-19, P = 6.2e-19
 Identities = 55/133 (41%), Positives = 80/133 (60%)

```

Query:      8 VHLFAGGCGGTVGAILTCPLEVVKTRLQSSS-VTLYISEVQLNTMAGA----SVNRVVSP 62
              VH AGG GG GA++TCP ++VKTRLQS + Y S+ +N G+ S+N V+
Sbjct:     54 VHFVAGGIGGMAGAVVTCFDFLVKTRLQSDIFLKAYSQA-VNISKGSTRPKSINYVIAQ 112

```


Query: 63 GP-----LHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCCKEKLNDVFD--P 115
 G L + + ++EG RSLF+GLGPNLVGV P+R+I F Y K+ F+
 Sbjct: 113 GTHFKETLGIIGNVYKQEGFRSLFKGLGPNLVGVIPARSINFFTYGTTKDMYAKAFNNGQ 172

Query: 116 DSTQVHMISAAMAG 129
 ++ +H+++AA AG
 Sbjct: 173 ETPMIHLMMAATAG 186

Score = 77 (11.6 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 25/88 (28%), Positives = 39/88 (44%)

Query: 3 QRDTLVHLFAGGCGGTGVAITCPLLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVSP 62
 Q ++HL A G A T P+ ++KTR VQL+ SV + +
 Sbjct: 172 QETPMIHLMAAATAGWATATATNPILIKTR-----VQLDKAGKTSVRQYKNS 219

Query: 63 GPLHCLKVILEKEGPRSLFRGLGPNLVG 90
 CLK ++ EG L++GL + +G
 Sbjct: 220 WD--CLKSVIRNEGFTGLYKGLSASYLG 245

Score = 71 (10.7 bits), Expect = 6.6e+00, P = 1.0e+00
 Identities = 28/91 (30%), Positives = 45/91 (49%)

Query: 12 AGGCGGTGVAITCPLLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVSPGPLHCLKVI 71
 + G V +I T P EVV+TRL+ + + N G R + G + KVI
 Sbjct: 294 SAGLAKFVASIATYPHEVVRTRLRQTP-----KEN---G---KRKYT-GLVQSFKVI 338

Query: 72 LEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY 102
 +++EG S++ GL P+L+ P+ I F +
 Sbjct: 339 IKEEGLFSMYSGLTPHLMRTVPNSIIMFGTW 369

Pedant information for DKFZphtes3_20m18, frame 1

Report for DKFZphtes3_20m18.1

[LENGTH] 132
 [MW] 13993.36
 [pI] 8.42
 [HOMOL] PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast (*Saccharomyces cerevisiae*) 7e-19
 [FUNCAT] 07.16 purine and pyrimidine transporters [*S. cerevisiae*, YBR192w] 3e-20
 [FUNCAT] 08.04 mitochondrial transport [*S. cerevisiae*, YBR192w] 3e-20
 [FUNCAT] 30.16 mitochondrial organization [*S. cerevisiae*, YBR192w] 3e-20
 [FUNCAT] 02.13 respiration [*S. cerevisiae*, YBR192w] 3e-20
 [FUNCAT] 01.05.07 carbohydrate transport [*S. cerevisiae*, YPR021c] 3e-10
 [FUNCAT] 07.07 sugar and carbohydrate transporters [*S. cerevisiae*, YPR021c] 3e-10
 [FUNCAT] 07.99 other transport facilitators [*S. cerevisiae*, YEL006w] 1e-09
 [FUNCAT] 01.07.10 transport of vitamins, cofactors, and prosthetic groups [*S. cerevisiae*, YEL006w] 3e-09
 [FUNCAT] 07.04.07 anion transporters (cl, so4, po4, etc.) [*S. cerevisiae*, YKL120w] 2e-08
 [FUNCAT] 01.03.19 nucleotide transport [*S. cerevisiae*, YPR011c] 3e-08
 [FUNCAT] 04.05.03 mRNA processing (splicing) [*S. cerevisiae*, YKR052c] 4e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [*S. cerevisiae*, YJR095w] 2e-07
 [FUNCAT] 01.01.07 amino-acid transport [*S. cerevisiae*, YOR130c] 5e-05
 [FUNCAT] 07.10 amino-acid transporters [*S. cerevisiae*, YOR130c] 5e-05
 [FUNCAT] 01.04.07 phosphate transport [*S. cerevisiae*, YJR077c] 7e-05
 [FUNCAT] 13.04 homeostasis of other ions [*S. cerevisiae*, YJR077c] 7e-05
 [BLOCKS] BL00215B Mitochondrial energy transfer proteins
 [BLOCKS] BL00215A Mitochondrial energy transfer proteins
 [PIRKW] duplication 6e-09
 [PIRKW] transmembrane protein 6e-09
 [PIRKW] mitochondrial inner membrane 4e-07
 [PIRKW] transport protein 5e-06
 [PIRKW] mitochondrion 7e-08
 [PIRKW] chloroplast 3e-08
 [SUPFAM] Btl protein 3e-08
 [SUPFAM] ADP,ATP carrier protein repeat homology 4e-09
 [SUPFAM] Caenorhabditis probable carrier protein c2 4e-09
 [SUPFAM] probable carrier protein YPR021c 6e-09
 [PROSITE] LEUCINE ZIPPER 1
 [PROSITE] MITOCH_CARRIER 1
 [PFAM] Mitochondrial carrier proteins
 [KW] Alpha_Beta

SEQ MSQRDTLVHLFAGGCGGTGVAITCPLLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV


```

PRD      cccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccccc
SEQ      SPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDFDPDSTQV
PRD      cccchhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccc
SEQ      HMISAAMAGMNV
PRD      chhhhhhhcccc

```

Prosites for DKFZphtes3_20m18.1

```

PS00029      27->49      LEUCINE ZIPPER      PDOC00029
PS00215      26->36      MITOCH_CARRIER      PDOC00189

```

Pfam for DKFZphtes3_20m18.1

```

HMM_NAME      Mitochondrial carrier proteins
HMM            *pFwkdfLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpR.....
               ++++++AGG +G + +++++P++++KTR+Q++ ++ + ++
Query          5  DTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSS-SVTLYISEVQLNTMA      52
HMM            .....YkGMIdCFRwiwkNEGWRGLWRGLGANVIRYIPqWaiRFGFY
               G+++C++ I+++EG+R+L+RGLG+N++++P +AI+F+ Y
Query          53 GASVNRVVSPPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY      102
HMM            EFMKeMFiDyfgeddnYmWfWmnYMaGs*
               +KE ++D F++ D+++++ + +MAG+
Query          103 SNCKEKLNDFDP-DSTQVHMISAAMAGM      130

```


DKFZphtes3_21d4

group: signal transduction

DKFZphtes3_21d4 encodes a novel 464 amino acid putative GTP exchanging factor related to RCC1.

RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator.

The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

similarity to RCC1-like G exchanging factor RLG

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="20"

Insert length: 2321 bp

Poly A stretch at pos. 2293, polyadenylation signal at pos. 2262

```
1 GGGTCACGCA AGATGGCGGC GCCCAGAGGC TGCTGAGGCG CGGAACGGAG
51 GATGGCGCTG GTGGCGTTGG TGGCTGGGGC TCGGCTGGGG CGGCGGCTGA
101 GCGGGCCGGG GCTGGGGCGA GGGCACTGGA CGGCGGCCAG GCGCTCCCGG
151 AGCCGGCCGG AAGCGGCAGA AGCCGAGGCG GAGGTGCCCG TGGTCCAGTA
201 CGTGGGGCGAG CGCGCTGCCC GCGCCGATCG CGTCTTCGTG TGGGGCTTCA
251 GCTTCTCGGG GCGCTGGGCG GTGCCTTCCT TTGTGGTGCC CAGCTCCGGG
301 CCCGGGGCCC GCGCCGGGCG CCGACCGCGC CGCAGGATCC AGCCCGTGCC
351 CTATCGCCTG GAGCTGGACC AAAAGATTTC ATCTGCTGCT TGCGGCTATG
401 GATTCACTCT GCTGTCTCTT AAGACTGCGG ATGTTACGAA AGTCTGGGGG
451 ATGGGACTCA ACAAAGATTG TCAGCTTGGA TTTCACAGGA GCCGGAAGA
501 TAAACGAGG GGCCTACGAGT ATGTGTTGGA GCCCTCAGCC GTCTCCCTGC
551 CTCTGGACAG ACCTCAGGAG ACACGGGTGC TGCAGGTCTC CTGCGGCCGA
601 GCTCACTCTC TTGTGTTGAC TGACAGGGAA GGAGTCTTCA GCATGGGAAA
651 CAATTCTTAT GGGCAATGTG GAAGAAAGGT GGTCGAAAAT GAAATTTACA
701 GTGAAAGTCA CAGAGTCCAC AGGATGCAGG ACTTCGATGG CCAGGTGGTC
751 CAGGTGCGCT GTGGTCAGGA TCATAGTCTG TTCCTGACGG ATAAAGGAGA
801 AGTCTATTCT TGTGGATGGG GTGCTGATGG GCAACAGAGT CTGGGTCACT
851 ACAATATCAC CAGCTCGCCC ACCAAGCTGG GTGGAGACCT GCGGGAGTG
901 AACGTTATCC AAGTTGCCAC CTACGGTGAT TGCTGCTTGG CCGTGTCCGC
951 CGACGGAGGA CTTTTGGTTT GGGGAAACTC GGAGTACCTG CAGCTGGCCT
1001 CTGTCACTGA CTCCACACAG GTGAATGTGC CCGCTGCTT ACACCTTCTCA
1051 GGAGTGGGGA AGGTGCGACA GGCTGCATGC GGTGGCACGG GCTGTGCACT
1101 GTTAAACGGA GAAGGACATG TTTTGTCTG GGGCTATGGA ATTCTTGGGA
1151 AAGGTCCAAA CCTAGTGGAA AGTGCCGTCC CTGAAATGAT TCCACCCACT
1201 CTCTTTGGCT TGACGGAGTT CAACCCAGAA ATCCAGGTTT CCCGCATCCG
1251 ATGTGGACTC AGCCACTTTG CTGCACTGAC CAACAAGGA GAGCTGTTTG
1301 TATGGGGCAA GAACATCCGA GGGTGCTTGG GAATCGGTCT CCTGGAGGAC
1351 CAGTATTTCC CATGGAGGCT GACGATGCCT GGGGAGCCTG TGGACGTGGC
1401 ATGTGGCGTG GACCACTAGG TGACCTTGGC CAAGTCATTC ATCTAAACCT
1451 CCCTCACCTG CTTGGGCGGC CCCGTCCCGG GAACCACTGG CACTCCTTGG
1501 CAGAGGCCAG CGCGTGGCCA GCGCCCGGGG GTTCTTGGAT GGTGGTGGCG
1551 GAGGACCCCTG CGTGCAGTGT GACGCTCTGT CCTGAATCCC TTAGCGGGTA
1601 CCTACCAGGA GGATCAGGGC AAGGTCCCTC TCCAGCTGCA GGTGAGGCCT
1651 GCGGAACCTA GCTTGGATGG CAGCCTTTGG TGGGCCGCTG TGGCCCGCAC
1701 GTCTCTGTTT TCTCCAAGTA ACATGCGACG GTGTCTGGTG TCACGTCTCG
1751 CCTGAGAAGC CCGTCTTAGG AAAGCTTAGC TTGAACACAG TGCTCGGGAG
1801 GTTCTGTGCT TGTCTGTCTT GGCAGTCTCT TGGTTTGTGT CTGGCCAAGG
1851 CCATGCGTGT GCCTCGGACC GAGCCCCAGC TTAGGCGAGG GAGTCAGGCT
1901 GGCTTCGGCC CTCGGTTTTC ATTACAGCCA CCCTGCTCAT GGCCCTTCCT
1951 GGCCGCGCTG CACACCGCAA GCTCGCTGGG GGGACACTAG AAGCACCGTG
2001 GCCTGGGATT CCATCTGGAG CTGTCCGAG GCACCAAGCC CAGCCTCCCA
2051 CCACGCTCAC TGCTTGGCTT GGAAGGTTA AGAAGCCCTT CAGGAAGAGA
2101 ATCGAGGCTA AGTTCCTCTG CGCCGAGGGC CCCGAGCATA TCCGCCAAGG
2151 CTCAGCTGCA GTGCCAGGCG GAGGAGGAAG ATCCAGAAAT TGTGAACAAT
2201 GTTTGATTTA GTAGCGTGAC TTGCCITTCC CTTTAAAAAC ATCTTTTACA
2251 AATCTGTCTT GGAATAAAGT CTATTTTCTG CCTTTTGGTT TTTAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry HS203358 from database EMBL:
human STS SHGC-31781.
Score = 1748, P = 1.1e-72, identities = 376/394

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 52 bp to 1443 bp; peptide length: 464
Category: similarity to known protein

```

1 MALVALVAGA RLGRLSGPG LGRGHWTAAR RSRSRREAAE AEAEPVVQY
51 VGERAARADR VFVWGFSGFSG ALGVPSFVVP SSGPGPRAGA RPRRRIQPVP
101 YRLELDQKIS SAACGYGFTL LSSKTADVTK VWGMGLNKDS QLGFRHSRDK
151 KTRGYEYVLE PSPVSLPLDR PQETRVLQVS CGRAHSLVLT DREGVFSMGN
201 NSYGGQGRKV VENEIYSESH RVHRMQDFDG QVVQVACGQD HSLFLTDKGE
251 VYSCGWGADG QTGLGHYNIT SSPTKLGGDL AGVNVIQVAT YGDCCLAVSA
301 DGGFLFGWNS EYLQLASVTD STQNVNPRCL HFGVGKVRQ AACGGTGCAV
351 LNAGEHVFWV GYGILGKGPV LVESAVPEMI PPTLFLGLEF NPEIQVSRIR
401 CGLSHFAALT NKGELEFVWGK NIRGCLGIGR LEDQYFPWRV TMPGEPVDVA
451 CGVDHMTLA KSFI

```

BLASTP hits

Entry CEW09G3_5 from database TREMBLNEW:
gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3
Score = 395, P = 9.3e-37, identities = 111/330, positives = 165/330

Entry Y032_HUMAN from database SWISSPROT:
HYPOTHETICAL PROTEIN KIAA0032.
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry B38919 from database PIR:
hypothetical protein 2 - human (fragment)
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry AF060219_1 from database TREMBLNEW:
product: "RCC1-like G exchanging factor RLG"; Homo sapiens RCC1-like G
exchanging factor RLG mRNA, complete cds.
Score = 273, P = 4.0e-21, identities = 84/262, positives = 124/262

Entry S71752 from database PIR:
giant protein p619 - human
Score = 282, P = 1.1e-19, identities = 86/287, positives = 144/287

Alert BLASTP hits for DKFZphtes3_21d4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_21d4, frame 1

Report for DKFZphtes3_21d4.1

```

[LENGTH]      464
[MW]           49997.08
[pI]           8.74
[HOMOL]        TREMBL:CEW09G3_5 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3 5e-34

[FUNCAT]       04.07 rna transport      [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
               [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       08.01 nuclear transport   [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
               cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.01.04 rrna processing   [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.03.03 trna processing   [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YGL097w] 2e-09

```



```

[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YAL020c] 4e-06
[BLOCKS] BL008701
[BLOCKS] BL00625B Regulator of chromosome condensation (RCC1) proteins
[BLOCKS] BL00625A Regulator of chromosome condensation (RCC1) proteins
[PIRKW] blocked amino end 3e-16
[PIRKW] nucleus 3e-16
[PIRKW] duplication 4e-08
[PIRKW] tandem repeat 3e-16
[PIRKW] DNA binding 3e-16
[PIRKW] mitosis 3e-16
[PIRKW] leucine zipper 3e-21
[SUPFAM] pheromone response pathway component SRM1 4e-08
[SUPFAM] WD repeat homology 3e-21
[PROSITE] MYRISTYL 7
[PROSITE] RCC1_2_2 2
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 5
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] GLYCOSAMINOGLYCAN 3
[PROSITE] PKC_PHOSPHO_SITE 7
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] Regulator of chromosome condensation (RCC1)
[KW] All_Beta
[KW] LOW_COMPLEXITY 13.58 %

```

```

SEQ MALVALVAGARLGRRLSGPGLGRGHWTAARRSRSRREAAEAEVPPVQYVGERAARADR
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccchhhhhhhhhheeeccccccccchhhhhhhhhhhhhhhhhhhhhceeeehhhhhhhhh

SEQ VFVWGFSGFSGALGVPSFVVPSSGPGPRAGARPRRRIQVPYRLELDQKISSAACGYGFTL
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD eeeccccccccccccccccccccccccccccccccccccchhhhhhhheeecccccccc

SEQ LSSKTADVTKVWGMGLNKDSQLGFHRSRKDKTRGYEYVLEPSPVSLPLDRPQETRVLQVS
SEG .....
PRD ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ CGRAHSLVLTDRGVFSMGNNSYGQCGRKVVENEIYSESHRVHRMQDFDQGVVQVACGQD
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HSLFLTDKGEVYSCGWGADGQTGLGHYNITSSPTKLGGDLAGVNVIVQVATYGDCCCLAVSA
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DGGLFGWGNSEYLQLASVTDSTQVNVPRCLHFSVGKVRQAACGGTGCAVLNNGEGHVFWV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GYGILGKGNLVEASVPEMIPPTLFLGLTEFNPEIQVSRIRCGLSHFAALTNKGELEFVWGK
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ NIRGCLGIGRLEDQYFPWRVTMPGEPVDVACGVDHMTLAKSFI
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_21d4.1

PS00001	200->204	ASN_GLYCOSYLATION	PDOC00001
PS00001	268->272	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	82->86	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	333->337	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	14->18	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	219->222	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	290->294	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006

PS00007	209->217	TYR_PHOSPHO_SITE	PDOC00007
PS00007	208->217	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	238->244	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	302->308	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00009	12->16	AMIDATION	PDOC00009
PS00009	206->210	AMIDATION	PDOC00009
PS00626	179->190	RCC1_2	PDOC00544
PS00626	235->246	RCC1_2	PDOC00544

Pfam for DKFZphtes3_21d4.1

HMM_NAME	Regulator of chromosome condensation (RCC1)		
HMM	*IAaGqHHTVCLTqDGRVYtWG*		
	+A GQ+H++ LT++G VY++G		
Query	235	VACGQDHSFLTDKGEVYSCG	255

DKFZphtes3_21j15

group: transcription factors

DKFZphtes3_21j15 encodes a novel 898 amino acid protein with similarity human NY-CO-33 protein.

NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

strong similarity to "NY-CO-33"

complete cDNA, complete cds, potential start at bp 27, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 4407 bp

Poly A stretch at pos. 4321, polyadenylation signal at pos. 4301

```

1 CGCTGCAGCA GGTGTCACAG AGCCGCATGC TCCCGGAGCC CAGCCTCTTC
51 AGCACCGTGC AGCTGTACCG GCAGAGCAGC AAGCTCTATG GCTCCATCTT
101 CACGGGGGCC AGCAAGTTCC GCTGTAAGGA CTGCAGCGCT GCCTACGACA
151 CCCTGGTGGG GTTGACAGTG CACATGAACG AGACGGGGCA TTACCGCGAC
201 GACAACCATG AGACCGATAA CAACAACCCC AAGCGCTGGT CCAAGCCTCG
251 CAAACGCTCC TTGCTGGAAA TGAAGGGGAA GGAAGACGCC CAGAAGGTGC
301 TGAAGTGCAT GTACTGTGGC CACTCCTTTG AGTCCCTGCA GGATTGTAGT
351 GTCCATATGA TCAAAACAAA AACTACCAA AAAGTGCCCT TGAAGGAACC
401 CGTCACTCTT GTCGCCGCCA AAATCATCCC TGCCACTCGG AAGAAAGCTT
451 CCCTGGAGCT GGAGCTCCCC AGCTCCCCAG ATTCCACAGG TGAACCCCCC
501 AAAGCCACCA TCTCAGACAC CAACGATGCA CTTCAGAAGA ACTCCAACCC
551 TTACATCAGC CCAAATAATC GGTACGGCCA CCAGAATGGG GCCAGCTATG
601 CATGGCATTG TGAGGCCCGG AAGTCGCAGA TCCTGAAAGT CATGGAGTGT
651 GGGAGCTCGC ATGACACCCT GCAGGAGCTC ACTGCCACA TGATGGTCAC
701 TGGCCACTTC ATCAAGGTCA CCAACTCTGC TATGAAAAGG GGAAGGCCCA
751 TTGTGGAGAC GCCTGTGACA CCTACCATCA CAACCCCTGT GGATGAGAAG
801 GTCCAGTCCG TGCCCTGGC AGCCACCACC TTCACGTCCC CCTCCAATAC
851 ACCTGCCAGC ATCTCCCAA AACTGAATGT GGAGGTCAAG AAGGAAGTCG
901 ACAAGGAGAA AGCGGTCACT GACGAGAAAC CTAAGCAAAA AGACAAGCCT
951 GGCGAAGAAG AGGAGAAGTG TGACATCTCT TCCAAATACC ATTACTTGAC
1001 TGAAAATGAC TTAGAAGAGA GTCCCAAGGG GGGCTTGAT ATCCTCAAAT
1051 CTTTGGAAAA CACAGTGACA TCCGCAATCA ACAAGGCCCA GAACGGCACT
1101 CCTAGCTGGG GGGGCTATCC CAGCATCCAT GCCGCTACC AACTTCCCAA
1151 CATGATGAAG TTGTCCCTGG GCTCGTCGGG GAAGAGCAGC CCCCTGAAAC
1201 CCATGTTTGG CAACAGTGAG ATTGTCTCCC CGACGAAAAA CCAGACCCTG
1251 GTCTCTTCCA CCAGCAGCCA GACGTCCCCC ATGCCCAAGA CAAACTTTCA
1301 TGCCATGGAG GAGCTGGTGA AAAAGGTGAC TGAGAAAGTT GCCAAAGTGG
1351 AGGAGAAGAT GAAGGAGCCG GATGGGAAGC TTTCCCGGCC CAAGCGGGCC
1401 ACTCCCTCCC CATGTAGCAG CGAAGTCGGG GAACCCATCA AGATGGAGGC
1451 ATCCAGCGAT GGGGGCTTCC GCAGCCAGGA GAACAGCCCC AGCCCCCGGC
1501 GGGATGGGTG CAAGGATGGG AGCCCCCTCG CTGAGCCGGT GGAGAATGGC
1551 AAGGAGCTGG TGAAGCCCTT AGCCAGCAGT TTGAGTGGCA GCACGGCCAT
1601 CATCACCAGC CACCCGCTGT AACAGCCTTT TGTTAACCTT TTGAGCGCCC
1651 TGCAGTCAGT CATGAACATT CACCTGGGCA AGGCCGCCAA GCCCTCCCTG
1701 CCTGCCCTGG ACCCCATGAG CATGCTTTTC AAGATGAGCA ACAGCCTGGC
1751 GGAGAAGGCT GCTGTGGCCA CCCCAGCGCC CCTGCAGTCC AAGAAGGCAG
1801 ACCACCTCGA CCGCTATTTC TACCACGTCA ACAACGACCA GCCCATAGAC
1851 TTGACAAAAA GGAAGAGTGA CAAAGGCTGC TCCTTGGGTT CAGTGCTTCT
1901 GTCACCCACG TCCACAGCCC CGGCAACCTC CTCATCCACG GTGACAACGG
1951 CAAAGACATC TGCCGTGCGT TCATTATGTT CAAACTCGCC GCTACGCGAG
2001 AATGCCTTGT CAGATATATC CGATATGCTG AAGAACTTGA CAGAGAGCCA
2051 CACGTCAAAA TCCTCCACTC CTTCCAGCAT CTCCGAGAAG TCTGACATTG
2101 ACCGGGGCCC TCTGGAGGAG GCTGAGGAGT CGACGCCCCG CCAGAAGAGG
2151 AAGGGCCGCC AGTCAAACGT GAACCCCCAG CACCTCCTGA TCCTCCAGGC
2201 CCAGTTTGCC GCCAGCCTCC GGCAGACCTC AGAAGGGAAG TACATCATGT
2251 CAGACCTGAG CCCCAGGAG CGGATGCATA TCTCCAGGTT CACCGGGCTG
2301 TCCATGACCA CCATCAGCCA CTGGCTGGCC AACGTGAAAT ACCAGCTTCG
2351 AAGGACAGGT GGAACAAAGT TCCTCAAAAA CTTGGACACT GGCCACCCCG
2401 TCTTCTTTTG TAACGATTGT GCGTCCCAAA TCAGGACTCC TTCCACGTAC
2451 ATCAGTCACC TAGAGTCACA CTTAGGCTTC CGGCTACGGG ACTTATCCAA
2501 ACTGTCCACC GAACAGATTA ACAGTCAGAT AGCACAAACC AAGTCACCGT
2551 CAGAAAAAAT GGTGACGTCC TCCCCGAGG AAGACCTGGG GACTTCCTAT
2601 CAGTGCAAACT TTTGCAATCG GACCTTGGCC AGCAAGCAGC CTGTTAAACT

```



```

2651 TCACCTTAGC AAAACACACG GGAATCTCC GGAAGACCAC CTTCTGTATG
2701 TCTCTGAGTT AGAGAAGCAG TAGCATTGTC TTTTGATAGA AAGGACTGCA
2751 GTTTGCTTTG AGGGAACTG TGGGAAGGCAC CTTCAGGCC CTTCTGACTT
2801 GTTGTCTTGC GCACATGTTT TTATTTTAAAC TGCAGAGAAT CACTCTGGGC
2851 TGGACTGTTT TGTATAACTG TACAGTGTTC AATAGAGGTG CATAATCAGC
2901 TGTGTGTTACT GGTAAAATAT GAAGGTAAAT ATGCAGTGGT AAGTGTGTTG
2951 AACTTTGTGT AAACGGGATT TAGTTGTGAG CATCCTCCCG ATGCTTCAAG
3001 CTGCATGCAT TAACAGACAG TTTAATTAAG CATTATAAC GGAATCAGGC
3051 ACACCTTTTC CACGAGACTC GAGTGTGCTG GCATTCTCA CCCTTTCATC
3101 TTTAGCCCTC TGAGTACTTT GAAGCACTTT TGCATTAATT TGGTTAAAAA
3151 ATAAAAATAA ATAATAATAA TGTATGAAGC TCTGTTTTT AAACCTCTTA
3201 CCAGCTTAGT TATAATGAAT AATATGAACC TCCATTATG CAGGTCTGCA
3251 GGGGTATAAC ACGCCTTGAA ATTTAAAGA ATATTATTT CACATTGAAA
3301 CATAGATGTA TATATTGTAT AGATTTCAGA CTCTCTTATG AAAAAAATG
3351 TGATTGTGGT TAAATGACCT TTTCTTGCA TTTATAGCAA CAGTGTTTTA
3401 TGCACCTGCT ATGCTCTGGG CATAAGCTGT GCCTATGTAT AGTGTATATT
3451 TCTTTTTTTC TTTTTTTTAA GGTCTATGGG TTTGTTTTT TACATGCAAA
3501 CATTGTAAAT TATACAGAAG ATACCACAGA TAGCATTAT AAAGTATACA
3551 GAAACATTAT CTGAAAGCAA AGTATGATAG TTTGTTTTGC TATACAGTAC
3601 ATCTATATTG ATAGAGGTTT ATGTTTAAAT TATACATATT TATTAGCATC
3651 ATATTGTCAT TTGTTTTGAG CAGTCTGAAT AAACGAGACC GGGAAAGACA
3701 TCCCTGGCAG GCATCAGAAC TATTTGTCAC ATGATTTTTA AAGGTATTTA
3751 TTAGAAATCA AAGAACAATC AAAATAAACT CAGTGCTCAA AGGGTTAAGT
3801 CTATTTGAAA AGGTAAAAAA AAAGAACAAA AAAAAAATAA GAACTTGATC
3851 TGTATTTCTT AAACATTGAT AAAGCCTTTA AAATGTTGT ACTGTAATAC
3901 TTTGCTTAAA AGTCATGAGG CATCTGTGA TCCAACCTCT TTCACTTATT
3951 TATAAGCCCT CTGGGTGCT ATTCCATATT GTAGGATGCC TTTCTATTTC
4001 AATTGGTAAC TTTCTGTTT GTTCTTCCTA ATTATTCTCC CAAGATCCCA
4051 CACTGCAGCT TTATCTTTAG GCTTATGAAA GGTAAACCCGT GGTACCAGGC
4101 TCTCCAGTG ATTCTGTCT TCTCCATTTT TGGCAGTTAA TTTGCAGAAG
4151 TAACTGACAG CTGACACCAT ATGAGAACCT TTGTATAAAA TATTGGCATG
4201 TAAACAGCAC AGACACCGTA ACACACTCTG TGCCCTGTTT GGTGTGTGAC
4251 AATGAAGCAC CATTATGTGA CTCTTCATAT AACCTTTTT TCTACGGCAG
4301 CATTAAAATT GTCTTTTTGC TATAAAAAAA AAAAAAATAA AAAAAAATAA
4351 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4401 AAAAAAATAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 2720 bp; peptide length: 898
 Category: strong similarity to known protein

```

1 MLPEPSLFST VQLYRQSSKL YGSIFTGASK FRCKDCSAAY DTLVELTVHM
51 NETGHYRDDN HETDNNNPKR WSKPRKRSLL EMEGKEDAQK VLKCMYCGHS
101 FESLQDLVSH MIKTKHYQKV PLKEPVTVA AKIIPATRK ASLELELPSS
151 PDSTGGTPKA TISDTNDALQ KNSNPYITPN NRYGHONGAS YAWHFEARKS
201 QILKCMCEGS SHDTLQELTA HMMVTGHFIK VTNSAMKKGK PIVETPVTPT
251 ITTLLDEKQV SVPLAATTFT SPSNTPASIS PKLNVEVKKE VDKEKAVTDE
301 KPKQKDKPGE EEEKCDISSK YHYLTENDLE ESPKGGLDIL KSENTVTSTA
351 INKAQNGTSP WGGYPSIHAA YQLPNMMKLS LGSSGKSTPL KPMFGNSEIV
401 SPTKNQTLVS PPSSQTSMP KTNFHAMEEL VKKVTEKVAK VEEKMKEPDG
451 KLSPPKRATP SPCSSEVGEP IKMEASSDGG FRSQENSPSP PRDGGCKDGS
501 LAEPVENGKE LVKPLASSLS GSTAIITDHP PEQPEVNPLS ALQSVMIHL
551 GKAAPSLPA LDPMSMLFKM SNSLAEKAAV ATPPLQSKK ADHLDRYFYH
601 VNNQPIDLT KGKSDKGC SL GSVLLSPTST APATSSSTVT TAKTSVVSF
651 MSNSPLRENA LSDISDMLKN LTESHTSKSS TPSSISEKSD IDGATLEAE
701 ESTPAQKRKG RQSNWNPOHL LILQAFQFAS LRQTSSEKYY MSDLSPOERM
751 HISRETGLSM TTISHWLANV KYQLRRTGGT KFLKNLDTGH PVFFCNDACS
801 QIRTPSTYIS HLESHLGFRL RDLKSLSTEQ INSQIAQTKS PSEKMTSSP
851 EEDLGTSYQC KLCNRTFASK HAVKLHLSKT HGKSPEDHLL YVSELEKQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_21j15, frame 3

TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds., N = 1, Score = 1039, P = 5.5e-105

PIR:A38437 probable homeotic protein tsh - fruit fly (Drosophila melanogaster), N = 3, Score = 158, P = 7.2e-09

TREMBL:CE33058_1 gene: "unc-89"; product: "UNC-89"; Caenorhabditis elegans UNC-89 (unc-89) gene, complete cds., N = 2, Score = 175, P = 3.3e-07

>TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.
Length = 687

HSPs:

Score = 1039 (155.9 bits), Expect = 5.5e-105, P = 5.5e-105
Identities = 244/504 (48%), Positives = 319/504 (63%)

Query: 170 QKNSNPYITPNNRYGHONGASYAWHFEARKSQILKMECGSSHDTLQELTAHMMVTGHFI 229
QK +NPY+TPNNRYG+QNGASY W FEARK+QILKMECGSSHDTLQ+LTAHMMVTGHF+
Sbjct: 14 QKAANPYVTPNNRYGYQNGASYTWQFEARKAQILKMECGSSHDTLQELTAHMMVTGHFL 73

Query: 230 KVTNSAMKKGKPIVETPVTPTITLLDEKVQSVPLAATFTS-PSNT----PASISPKLN 284
KVT SA KKGK +V PV ++EK+QS+PL TT T P+++ P S +
Sbjct: 74 KVTTASAKKGKQLVLDPV-----VEEKIQSIPLPPTHTRLPASSIKKQPDSPAGSTT 126

Query: 285 VEVKKEVDKEKA-VTDEKPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGLDILKSL 343
E KKE +KEK V + K K++ + EK + S+ Y YL E DL++SPKGGLDILKSL
Sbjct: 127 SEEKKEPEKEKPPVAGDAEKIKEESEDSELEFEPSTLYPYLREEDLDDSPKGGLDILKSL 186

Query: 344 ENTVTSAINKAQNGTPSWGYPYPSIHAAYQLPMMKLSLSSGKSTPLKPMF-GNSEIVSP 402
ENTV++AI+KAQNG PSWGGYPYPSIHAAYQLP +K L ++ +S ++P + G + +S
Sbjct: 187 ENTVSTAISKAQNGAPSWGYPYPSIHAAYQLPGTVK-PLPAVQSVQVQPSYAGGVKSLSS 245

Query: 403 TKNQTLVSPFSSQTSMPKTNFHAMVELVKKVTERV-AKVEEKMKEPDGKLSPPKRATPS 461
++ L+ P S T P K+N AMEELV+KVT KV K EE+ E + K S K A S
Sbjct: 246 AEHNALLHSPGSLTPPHKSNVSAAMEELVEKVTGKVNKKEERPEKE-KSSLAKAA--S 302

Query: 462 PCSSEVGEPKMEASSDGGFRSQENSPPPRDGCKDGSPLAEPVENGKELVKPLASSLSG 521
P + E + K E S + Q+ P K PL NG E +K ++
Sbjct: 303 PIAKENKDFPKTEEVSG---KPKKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCN 359

Query: 522 STAITDHPPEQPFVNPLSALQSVNMNIHLGKAAKPSLPALDPMMSMLFKMSNSLAEKAAVA 581
+ II DH PE F+NPLSALQS+MN HLGK +KP P+LDP++ML+K+SNS+ +K
Sbjct: 360 NLGIIMDHSPGPSLTPPHKSNVSAAMEELVEKVTGKVNKKEERPEKEKSSSLAKAASPIAK 419

Query: 582 TPPPLQSKKADHLDRYFYHVNNQPIDLTGKGS DK-GCSLGSVLLSPTSTAPATSSSTVT 640
P K+AD +DRY+Y N+DQPIDLTG K+ S+ + SP + S +
Sbjct: 420 ATPV---KQADAIDRYYYE-NSDQPIDLTGSKNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 641 TAKTSAVVSFMSN-SPLRENALS DISDMLKNLTE 673
T + S S + E + +D S + L E
Sbjct: 476 KNLTGRLTPKSSTPSTVSEKSDADGSSFEALDE 509

Score = 865 (129.8 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 211/434 (48%), Positives = 268/434 (61%)

Query: 447 EPDGKLSPPKRATPSPCSSEVG--EPIKMEASSDGGFRSQENSPPPRDG-CKDGSPLAE 503
E + L P TP P S V E + + + + +E P + K SP+A+
Sbjct: 247 EHNALLHSPGSLTPPHKSNVSAAMEELVEKVTGKVNKKEERPEKEKSSSLAKAASPIAK 306

Query: 504 -----P-VE--NGKELVK-PLASSLSGSTAITD-HPPE--QPFVNPLSALQSVNMNIHLG 551
P E +GK K P A + D H P +P ++ + + I +
Sbjct: 307 ENKDFPKTEEVSGKPKQKGPETWEAKKEGPLDVHTPNGTEPLKAKVTNGCNLGIIMD 366

Query: 552 KAAKPSLPALDPMMSMLFKMSNSLAEKAAVATPPPLQSKKADHLDRYFYHVNN---DQPID 608
+ +PS ++P+S L + N+ K + P L D L Y ++N D+P+
Sbjct: 367 HSPGPSF--INPLSALQSIMNTHLGKVKPVSPSL-----DPL-AMLYKISNSMLDKPV- 417

Query: 609 LTRGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENALS DISDML 668
K S P + + S+V ++ SPLRE+AL DISDM+
Sbjct: 418 -YPATPVKQADAIDRYYYENS DQPIDLTGSKNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 669 KNLTESHTSKSSTPSSISEKSDIDGATLEEA-EESTPAQKRKGROSNWNPOHLLILQAQF 727
KNLT T KSSTPS++SEKSD DG++ EEA +E +P KRKGROSNWNPOHLLILQAQF

Sbjct: 476 KNLGTGRLTPKSSTPSTVSEKSDADGSSFEELDELSPVHKKRGRQSNWNPQHLLILOAQF 535

Query: 728 AASLRQTSEGKYMISDLSPQERMHISRFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 787
A+SLR+T+EGKYMISDL PQER+HIS+FTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD

Sbjct: 536 ASSLRETTEGKYMISDLGPQERVHISKFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 595

Query: 788 TGHVPVFFCNDCASQIRTPSTYISHLESHLGFLRLDLSKLSTEQINSQIAQTKSPSEKMV- 846
TGHVPVFFCNDCASQ RT STYISHLE+HLGF L+DLSKL QI Q +K + K +

Sbjct: 596 TGHVPVFFCNDCASQFRTASTYISHLETHLGFLSLKDLKSLPLNQIQEQNVSKVLTNKTG 655

Query: 847 -TSSPEEDLGTSYQCKLCNRTFASK 870
+ EEDLG+++QCKLCNRTFA +

Sbjct: 656 PLGATEEDLGSTFQCKLCNRTFAKQ 680

Score = 98 (14.7 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 32/95 (33%), Positives = 47/95 (49%)

Query: 90 KVLKCMYCGHSFESLQDLSVHMIKTKHYQKVP-----KEPVT-PVAAKIIPATRKAS 142
++LKCM CG S ++LQ L+ HM+ T H+ KV K+ V PV + I + +

Sbjct: 45 QILKCMCGSSHDTLQQLTAHMMVTGHFLKVTTSASKKGKQLVLDPVVEEKIQSIPLPPT 104

Query: 143 LELELPSS-----PDSTGGTPKATISDTNDALQKNSNP 175
LP+S PDS G+ T S+ +K P

Sbjct: 105 THTRLPASSIKKQPDSPAGS---TTSEEKKEPEKEKPP 139

Score = 81 (12.2 bits), Expect = 4.6e-93, P = 4.6e-93
Identities = 13/29 (44%), Positives = 20/29 (68%)

Query: 28 ASKFRCKDCSAAYDTLVELTVHMMNETGHY 56
A +C +C +++DTL +LT HM TGH+

Sbjct: 44 AQILKCMCGSSHDTLQQLTAHMMVTGHF 72

Pedant information for DKFZphtes3_21j15, frame 3

Report for DKFZphtes3_21j15.3

[LENGTH] 898
[MW] 98486.72
[pI] 8.61
[HOMOL] TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens
antigen NY-CO-33 (NY-CO-33) mRNA, complete cds. 0.0
[BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins
[PIRKW] zinc finger 1e-06
[PIRKW] DNA binding 1e-06
[PIRKW] transcription regulation 1e-06
[PROSITE] MYRISTYL 9
[PROSITE] ZINC_FINGER_C2H2 4
[PROSITE] CAMP_PHOSPHO_SITE 5
[PROSITE] CK2_PHOSPHO_SITE 19
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 4
[PFAM] Zinc finger, C2H2 type
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 11.36 %

SEQ MLPEPSLFSTVQLYRQSSKLYGSIFTGASKFRCKDCSAAYDTLVELTVHMMNETGHYRDDN
SEG
PRD ccc

SEQ HETDNNNPKRWSKPRKRSLLEMEGKEDAQVLKCMYCGHSFESLQDLSVHMIKTKHYQKV
SEG
PRD ccc

SEQ PLKEPVT-PVAAKIIPATRKASLELELPSSPDSTGGTPKATISDTNDALQKNSNPYITPN
SEG
PRD ecc

SEQ NRYGHQNGASYAWHFEARKSQILKCMCGSSHDTLQELTAHMMVTGHFIKVTNSAMKKGK
SEG
PRD ccc

SEQ PIVETPVTPTITLLDEKVQSVPLAATFTSPSNTPASISPKLNVEVKKEVDKEKAVTDE
SEG xxx
PRD ccc

SEQ KPKQKDKPGEEEKCDISSKYHYLTENDLEESPKGGLDILKSLNTVTSAINKAQNGTPS

Prosites for DKFZphtes3_21j15.3

713

PS00006	810->814	CK2_PHOSPHO_SITE	PDOC00006
PS00006	840->844	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	884->888	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	732->740	TYR_PHOSPHO_SITE	PDOC00007
PS00007	883->892	TYR_PHOSPHO_SITE	PDOC00007
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	156->162	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	617->623	MYRISTYL	PDOC00008
PS00008	757->763	MYRISTYL	PDOC00008
PS00028	795->816	ZINC_FINGER_C2H2	PDOC00028
PS00028	860->882	ZINC_FINGER_C2H2	PDOC00028
PS00028	33->56	ZINC_FINGER_C2H2	PDOC00028
PS00028	94->117	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKFZphtes3_21j15.3

HMM_NAME Zinc finger, C2H2 type

HMM *CpwPDCgKtFrrwsNLrRHMR..T.H*
 C++ C ++ + +L+ HM+ H

Query 33 CKD--CSAAYDTLVELTVHMMNET-GH 55

26.69 (bits) f: 94 t: 116 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNLrRHMR..T.H*
 C + CG +F + +L HM+ H
 dkfzphes3 94 CMY--CGHSFESLQDLSVHMIKT-KH 116

Query f: 795 t: 815 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

HMM *CpwPDCgKtFrrwsNLrRHMRTH*
 C++ C R++S+++ H+ +H
 Query 795 CND--CASQIRTPSTYISHLESH 815

27.12 (bits) f: 860 t: 881 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNLrRHMR.T.H*
 C+ C++TF +++ + H+ H
 dkfzphes3 860 CKL--CNRTFASKHAVKLHLSK-TH 881

DKFZphtes3_21116

group: intracellular transport and trafficking

DKFZphtes3_21116 encodes a novel 66 amino acid protein nearly identical to rat ribosome attached membrane protein 4 (ramp4).

The novel protein seems to be the human orthologue of rat ramp 4. Ramp4 is involved in the regulation of translocation of proteins into endoplasmic reticulum, e.g. of the MHC class II associated invariant (gamma) chain.

The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

identical to rat ribosome attached membrane protein 4

ORF Bp 316-513 (66 aa) see BLASTX

Sequenced by LMU

Locus: unknown

Insert length: 2488 bp

Poly A stretch at pos. 2464, polyadenylation signal at pos. 2442

```
1 CTTCCTCTTT CACTCCGCGC TCACGGCGGC GGCCAAAGCG GCGGCGACGG
51 CGGCGCGAGA ACGACCCGGC GGCCAGTTCT CTTCCTCTCG CGCACCTGCC
101 CCGCTCGGTC AGTCAGTCGG CGGCCGCGCG CCGGCTTGTC CTCAGACCTC
151 GCGCTTGGCG CGCCCAGGCC CAGCGGCCGT AGCTAGCGTC TGGCCTGAGA
201 ACCTCGGCGC TCCGGCGGCG CGGGCACCCAC GAGCCGAGCC TCGCAGCGGC
251 TCCAGAGGAG GCAGGCGAGT GAGCGAGTCC GAGGGGTGGC CGGGGCAGGT
301 GGTGGCGCCG CGAAGATGGT CGCCAAGCAA AGGATCCGTA TGGCCAACGA
351 GAAGCACAGC AAGAACATCA CCCAGCGCGG CAACGTCGCC AAGACCTCGA
401 GAAATGCCCC CGAAGAGAAG GCGTCTGTAG GACCCTGGTT ATTGGCTCTC
451 TTCATTTTGG TTGTCTGTGG TTCTGCAATT TTCCAGATTA TTCAAAGTAT
501 CAGGATGGGC ATGTGAAGTG ACTGACCTTA AGATGTTTCC ATTCTCCTGT
551 GAATTTTAAC TTGAACATCAT TCCTGATGTT TGATACCCGT GTTGAAACAA
601 ATTCAGTAAA GCATCCTGCC TCAGAATGAC TTTCCATCA TGCTTCATGT
651 GTCATTCCAA GGTTCCTTCA TGAGTCATTC CAAGTTTCT AGTCCATACC
701 ACAGTGCCCT GCAAAAAACA CCACATGAAT AAAGCAATAA AATTGGATTG
751 TTAAGATACA GTAGTGGACC CTACTTATTC AGTCAATTAA GAGTAAGTTT
801 TTTTATGTGG TTATTAACAC AGTATGAACA ATTAGTCTAA CTCTGCATAG
851 ACAGGGTCTA GATTTTGTGA ACCCAAATGT ATAAGTCGAG TTAGCTTAAA
901 TTACAATTTG AAGTCTGTGG GTTTTATAT AGCTAGGCAC TTTATTACTC
951 TTTTGAAGTG AAAGCACACT CCCTTATAGG TTCATGTAAC TGTCCTGTAA
1001 TAAGGTGCTT ATAAATGGAA CAACTACACA GCCTAGTTT GCCACAACCT
1051 TTAGCATCTA AAAAGTTTAA AAAGCTTCTA AATGTCTAAT ATAAAGGGAG
1101 ATGCTTATAG CCACAACATC TATTTTACCA ATATTGTTTC CATTACACTA
1151 CCTTGGATTG TGCATGAGTG AGTATAGTAA CCCAAGATGC CATAAAAAAA
1201 AACTTGTATG TTTCTGACT TAATTAGTTA CTGTGGTTTC ACTAAAAGCT
1251 ACCGTGGTGG AGTGAAGTCA GTCAGGGAAG GTTTGTTTAT GTTACATTTA
1301 TTTACCAGA ACTATTTTAA TATATCAAAG GGGTTTACTA TGCCAAACAA
1351 AATTCTAGGG AAAAAATCTG CTAAAAATGG ATGCCTCATC AGAACATGCT
1401 GTTGAGTCCA ATGTGCCATA AGACATTTTA GCATGTTAAA TAGCACTTTT
1451 AATAGCAAAA AAAGGCACAT CAACTGCGAA GTTATCCTTA GTTTGCAAAAT
1501 GCTTTTCTA GATTAATGAT TTTTCAATCA TTAGGGTACT AGACACATCA
1551 GCCTAAAGTG GCATCTGGAA TTGAATGGAT TTACTGATAA TGATCAGTCT
1601 TTAGTCTTCC CTTTGTATA TGACTTTATA GGTATGATT GATCAAAATT
1651 ACGTTTTACT AATGGTAAGC GTGAGGTGTA TAGGGCAGGT TTTGGGTTTT
1701 CTAGTACTGT TGAAAATCTG AAGTATTGGC TATTTGTATA CTTAGCCATA
1751 ACTTGGTGAA AAAAAACCTG AGCAGTGTCT ATGTATTAAT GCGTTGGAAA
1801 GAAAGCTGCT TGTGTTTGCT TTGTTAATTG CCTCAGGATA TTTCTTTTAA
1851 AATAAGCTGT TTTAAGAGGA ACAGAAGGGA AATCTGCTAC CTAGTCTATA
1901 CACAGCGTGA ACCTCACAGG GGGCTTCTGA TACCCTCAAA CATGGAGAAC
1951 AGTAAGGGAG CAGAGTGGTT AAGGACTTTC AGGAACTTAA CTATTCTGGA
2001 ATAAGGAATG AATCAACTGA CCTTGGGCCA GCAGGTTTTT AACTAAATTG
2051 TTAAGTGGCT TTCTCACCCA GTTAATCAGT CTCTGACTT GTTCCCTTTT
2101 TTGAAACAAG TGTCTTGGTT AACTAATCTT GTTTTATGGT TGTGCTAAAT
2151 TCATAGCAGG TGCCCTATTTC TTTGCTTTTA GTCAAACCAT TCCATATCAG
2201 AATTTTCCTT GGTTTACTAT AGATATTGGT CTTTAAGTTG TTGTTTGTGT
2251 TTTTAAATGT ACAATGTTCT GATAAATTG ACTGTTAAAT TGCTATAGCT
2301 AGCAATCATT TTACATATGT AAAAAATTGC ATTCCCTTTG TATTTCATGT
2351 GTAATTCACC AATTAAGTGC AGTTTATATT CAGGTTGGAT TATGCATGTT
2401 TAGGTAACGC AAAGCTGTGT CTTACTTGAT TTATTCTTTA AAAATAAAGT
2451 TCCCTGAATA TTTGAAAAAA AAAAAAATAA AAAAAAATAA
```


BLAST Results

Entry HSCDN13 from database EMBL:
H.sapiens (TL5) mRNA from LNCaP cell line
Score = 1075, P = 5.8e-41, identities = 219/221

Entry AF100470_1 from database TREMBLNEW:
gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds.
Score = 331, P = 3.9e-28, identities = 66/66, positives = 66/66, frame +1

Entry HSG19910 from database EMBL:
human STS A002B48.
Score = 530, P = 2.1e-17, identities = 108/109

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 316 bp to 513 bp; peptide length: 66
Category: strong similarity to known protein
Classification: Intracellular transport and traffic

1 MVAKQIRIMA NEKHSKNITQ RGNVAKTSRN APEEKASVGP WLLALFIFVY
51 CGSAIFQIIQ SIRMGM

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21116, frame 1

TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4, N = 1, Score = 331, P = 6.2e-30

TREMBL:AF100470_1 gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds., N = 1, Score = 331, P = 6.2e-30

>TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4
Length = 75

HSPs:

Score = 331 (49.7 bits), Expect = 6.2e-30, P = 6.2e-30
Identities = 66/66 (100%), Positives = 66/66 (100%)

Query: 1 MVAKQIRIMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ 60
MVAKQIRIMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ
Sbjct: 10 MVAKQIRIMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ 69

Query: 61 SIRMGM 66
SIRMGM

Sbjct: 70 SIRMGM 75

No Pedant data available

DKFZphtes3_2ln23

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with strong similarity to rat 7acomp protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to rat 7acomp protein

on genomic level encoded by AF107885

Sequenced by LMU

Locus: /map="14q24.3"

Insert length: 3122 bp

Poly A stretch at pos. 3070, polyadenylation signal at pos. 3045

```
1 GGAAACCTC GTGGGCTCAG CCCGGGAGAA AGGGCCAGGG AAGTTGGGTG
51 GTTCTGTGCT TGGTCTGTCA ATGGAGGAGA TCAAAGTTTT ACGAAGGGTG
101 AAGGAGGAGA ATGATCGGCG AGGTGGATTT ATTCGCATAT TTCTACATC
151 TGAGACATGG GAAATATATG GGTCTACCT CGAGCATAAG ACCTCAATGA
201 ACTATATGCT GGCAACACGC CTCTTCCAGG ACAGGGGAAA CCAAGAAGA
251 AGCTTATTGA CAGGAAGAAC ACGAATGACT GCTGATGGAG CGCCAGAATT
301 GAAGATAGAG AGTCTGAATT CAAAGGCCAA GCTGCATGCT GCACTTTACG
351 AGAGGAAGCT CCTGTCTCTG GAGGTGCGAA AACGTAGACG ACGGAGTAGC
401 AGATTGAGGG CAATGAGGCC AAAATACCCA GTGATTACCC AACCAGCTGA
451 AATGAATGTT AAAACTGAGA CAGAGAGTGA AGAGGAGGAA GAAGTCGCAT
501 TAGATAATGA AGATGAAGAA CAGGAGGCTT CCCAGGAGGA GTCTGCAGGA
551 TTTCTTAGAG AAAATCAAGC CAAATATACA CCTCATTTGA CAGCTTTGGT
601 AGAAAATACA CCCAAGAAA ATTCCATGAA AGTTCGTGAA TGGAAATAATA
651 AAGGTGGACA CTGCTGCAAA CTTGAGACTC AGGAGCTAGA GCCTAAATTT
701 AACCTGATGC AGATTCTTCA AGATAATGGC AATCTTAGCA AAATGCAGGC
751 CCGAATAGCA TTCTCTGCCT ATCTCCAGCA TGTCAAATT CGCCTGATGA
801 AAGACAGTGG CGGTGAGACG TTCAGTGCCA GTTGGGCTGC CAAAGAGGAT
851 GAACAGATGG AGCTGGTTGT TCGTTTCCTC AAGCGAGCAT CAAATAACCT
901 CCAGCATTCA CTGAGGATGG TATTACCCAG TCGACGATTG GCACTTCTGG
951 AACGCAGAAG AATCCTGGCC CACCAGCTGG GTGACTTTAT CATTGTATAC
1001 AACCAAGAAA CAGAACAAT GGCTGAAAAG AAATCAAAGA AGAAAGTTGA
1051 GGAAGAAGAG GAAGATGGGG TGAATATGGA AAACCTTCAG GAGTTCATCA
1101 GACAAGCAAG TGAGGCTGAA CTGGAGGAGG TGTGACTTT TATACCCAA
1151 AAGAACAAGT CTGCTAGTGT CTTCCTGGGG ACTCACTCTA AAATTTCTAA
1201 GAACAACAAC AATTATTCTG ATAGTGGGGC AAAAGGTGAT CACCCTGAGA
1251 CTATAATGGA AGAAGTGAAA ATAAAGCCAC CTAACACGCA ACAGACGACA
1301 GAAATTCATT CTGATAAATT ATCTCGATT ACCACTTCAG CAGAAAAAGA
1351 GGCAAAATTA GTTTATAGCA ATTCTCTCTC TGGTCTTACT GCCTACTCTG
1401 AGAAAATTCC CAACACCCAT TTGTCATCTG TTACAACCTC TGACCTCTCT
1451 CCAGGGCCCT GCCACCATTC TTCTTTATCT CAAATTCCTT CAGCTATCCC
1501 CAGCATGCTC CACCAGCCAA CAATTTTACT GAACACAGTC TCTGCCAGTG
1551 CTCTCCCTG CCTACATCCC GGGGCACAGA ACATCCCAAG CCCTACTGGC
1601 CTGCCACGCT GTCGATCAGG AAGTCACACC ATTGGTCCCT TTTCTTCTCT
1651 CCAAAGTGCT GCACACATCT ATAGCCAGAA ACTGTCTCGT CCCTCTTCAG
1701 CAAAGGCAGG ATCGTGCTAT CTAACAAGC ATCATTGAGG AATAGCCAAA
1751 ACACAAAAAG AGGGAGAAGA TGCTTCTTTA TATAGCAAAC GGTACAACCA
1801 AAGTATGGTT ACAGCTGAAC TTCAGCGGCT AGCTGAGAAG CAGGCAGCGA
1851 GACAGTATTC TCCATCCAGC CACATCAACC TCCTCACCCA ACAGGTAACA
1901 AACCTGAATT TGGCAACTGG CATCATAAAC AGAAGCAGTG CTTAGCTTCC
1951 CCAACCCCTC CGACCCATCA TCAGTCTTAG TGGCCCGACA TGGTCTACAC
2001 AGTCAGACCC CCAAGCTCCC GAGAATCACT CCAGCTCTCC TGGAAAGCAGG
2051 AGCCTGCAGA CAGGGGGATT TGCCTGGGAA GGAGAAGTAG AAAACAACGT
2101 GTACAGCCAG GCTACAGGGG TGGTCCCCCA GCACAAGTAT CACCCACAG
2151 CAGGCAGCTA TCAGCTTCAA TTTGCCCTGC AGCAACTTGA ACAACAAAA
2201 CTTCACTCCC GGCAGCTCCT GGACCAAGAG CGAGCCCGGC ACCAGGCAAT
2251 CTTTGGCAGC CAGACACTAC CTAACCTCAA TTTATGGACA ATGAATAATG
2301 GTGCAGGTTG TAGAATTTC AGTGCCACAG CTAGTGGCCA GAAGCCAACC
2351 ACTCTGCCAC AAAAAGTGGT ACCACCTCCA AGTTCTTGCG CCTCCTGGT
2401 TCCCAACCC CCACCCAACC ACGAACAGT GCTCAGAAAG GCAACATCCC
2451 AGAAAGCTTC CAATACCCGC TTCAGATCCT CCTTCAAAA CTATTGTGG
2501 TATTTCTTCC AAGCAGTCAG CTGAACCTGAG GACGACAGCC TACAACAAC
2551 TACATGCATC TGAAGTGTCT CTTGTAATAT AGCTTTTTTC AGAGCCAGAA
2601 TCATACTCTC CAGGAATAT GGAGAAAGAA ACCTGAGGAG ATTGAAGTTT
2651 GCCAGGCACA AGGGCAAAAC TCAGACTGAA TGAATTTGAA AGGGTGGGGC
2701 CAAAGATGTT GTAACCTGGG AGACTTCTCT GAAGAAAGAA AACTGTTTAA
```



```

2751 GAAACACAGA CTGAACGCA GTACTTTTCC TTAAATAGCT GAGATGACCT
2801 TCTTTACCTT GGGCTTAGGT GATTCTCATC AGGGTGACCT CAGTGGAAGT
2851 TGGTGGTAAC GACTGTTCTG TGTCAGCACC CAGGACAGTG GTGTCTGTTA
2901 AGGCTGCCAG GGATTAGCAG GGAGGAAAGC CATCAGGACT GGGTAGCCTG
2951 GTAGACACAA ATCCCAATTA ATGTTACCTG AACATGTGGT GAGGTCAGCC
3001 GTATGATGAA AGATGTTTAA GAGATTAATG TCAGAAGAAT ATGAAAATAA
3051 ACACCGGCTT AAAAAATGTT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3101 AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry AF107885 from database EMBL:
Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes.
Score = 3042, P = 3.0e-219, identities = 610/612
5 exons matching 1893-3070

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 71 bp to 2521 bp; peptide length: 817
Category: strong similarity to known protein

```

1 MEEIKVLRV KEENDRRGGF IRIFPTSETW EIYGSYLEHK TSMNYMLATR
51 LFQDRGNPRR SLTGTRTRMT ADGAPELKIE SLNSKAKLHA ALYERKLLSL
101 EVRKRRRRSS RLRAMRPKYP VITQPAEMNV KTETESEEEE EVALDNEDEE
151 QEASQESAG FLRENQAKYT PSLTALVENT PKENSMKVRE WNNKGGHCKC
201 LETQELEPKF NLMQILQDNG NLSKMQARIA FSAYLQHVQI RLMKDSGGQT
251 FSASWAAKED EQMELVVRFL KRASNLIQHS LRMVLPSSRL ALLERRRILA
301 HQLGDFIIVY NKETEQAEMK KSKKKVEEEE EDGVNMFNFQ EFIRQASEAE
351 LEEVLTFTYT KNKSASVFLG THSKISKNNN NYSDSGAKGD HPETIMEEVK
401 IKPPKQQQTT EIHSKLSRF TTSAEKEAKL VYSNSSSGPT ATLQKIPNTH
451 LSSVTTSDLS PGPCHHSSLS QIPSAIPSMP HQPTILLNTV SASASPCLHP
501 GAQNIPSPPTG LPRCRSGSHT IGPFSFQSA AHYISQKLSR PSSAKAGSCY
551 LNKHHSGIAK TQKEGEDASL YSKRYNQSMV TAELOQLAEK QAARQYSPSS
601 HINLLTQQVT NLNLATGIIN RSSASAPPTL RPIISPSGPT WSTQSDPQAP
651 ENHSSSPGSR SLQTGGFAWE GEVENNVYSQ ATGVVPQHKY HPTAGSYQLQ
701 FALQQLQEQK LQSRQLDQS RARHQAIFGS QTLPNNSLWT MNNAGACRIS
751 SATASGQKPT TLPQKVPPPP SSCASLVFKP PPNHEQVLRR ATSQKASNTR
801 FRSSFQNYLW YFFQAVS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21n23, frame 2

TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA, complete cds., N = 1, Score = 1845, P = 2.2e-190

TREMBL:AF107885_3 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 443, P = 5.3e-41

TREMBL:AF107885_4 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 265, P = 8.2e-22

>TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA, complete cds.
Length = 436

HSPs:

Score = 1845 (276.8 bits), Expect = 2.2e-190, P = 2.2e-190
Identities = 369/435 (84%), Positives = 395/435 (90%)

Query:	115	MRPKYPVITQPAEMNVKTTETESEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLT	174
		MRPKYPVIT PAEMN+KTETESEEEEV LDNEDEEQEASQESAG L ENQAKYTPSLT	
Sbjct:	1	MRPKYPVITLPAEMNIKTETESEEEEVGLDNEDEEQEASQESAGSLAENQAKYTPSLT	60
Query:	175	ALVENTPKENSMKVREWNNGGHCCKLETQELEPKFNLMOILQDNGNLSKMQARIAFSAY	234
		+VEN+P+EN+MKV EW NKG CCK+ETQE E KFNLMQILQDNGNLSK+QAR+AFSAY	
Sbjct:	61	VIVENSPRENAMKVAEWTNKGESCCKIETQEPESKFNLMQILQDNGNLSKQARLAFSAY	120
Query:	235	LQHVQIRLMDKSDGGQTFSASWAAKEDEQMELVVRFLLKRASNQLQHSLRMVLPSRRLALLE	294
		LQHVQ+RL KDSGGQT S SWAAKEDEQMELVVRFLLKRAS+NLQHSLRMVLPSRRLALLE	
Sbjct:	121	LQHVQVRLTKDSGGQTLSPSWAAKEDEQMELVVRFLLKRASNQLQHSLRMVLPSRRLALLE	180
Query:	295	RRRILAHQLGDFIIVYNKETEQAEMAKSKKKVEEEEEEDGVNMENFQEFIRQASEAELEEV	354
		RRRILAHQLGDFI+VYNKETEQAEMAKSKKK+EEEEEDGVN E+QEFIRQASEAELEEV	
Sbjct:	181	RRRILAHQLGDFIIVYNKETEQAEMAKSKKKLEEEEEEDGVNAESFQEFIRQASEAELEEV	240
Query:	355	LTFYTTQKNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKKIPKQOQTTEIHS	414
		LTFYTTQKNKSASVFLGTHSK SKN+++YSDSGAKGDHPETI+EVKKI PKQOQ TEIHS	
Sbjct:	241	LTFYTTQKNKSASVFLGTHSKSKNNSSYSDSGAKGDHPETI+QEVKKI PKQOQOQTEIHS	299
Query:	415	DKLSRFTTSAEKEAKLVYNSNSS--GPTALT-QKIPNTHLSSV-TTSDLSGPGCHHSSL	470
		DKLSRFTTSA KEAKLVY+N SS GP A L Q++P+THLSS+ TTS LS GP HHSSL	
Sbjct:	300	DKLSRFTTSAGKEAKLVYTNCSFSGPAAVLLQLRPLSTHLSSIIITSTLSSGPGHHSSL	359
Query:	471	QIPSAIPSMPHQPTILLNTVSASASPLCHPGAQNIPTGLPCRSGSHTIGPFSSFQSA	530
		QI AIPSMPHQ +LLN V SASP +HPG N+ SP GLPCRSGS+TIGPFSSFQSA	
Sbjct:	360	QISPAIPSMPHQSALLLNVPVDSASPPVHPGTENV-SPAGLPCRSGSYTIGPFSSFQSA	418
Query:	531	AHIYSQKLSRPSSAKAG	547
		AHIYSQKLSRPSSAKAG	
Sbjct:	419	AHIYSQKLSRPSSAKAG	435

Pedant information for DKFZphtes3_21n23, frame 2

Report for DKFZphtes3_21n23.2

```
[LENGTH]          817
[MW]              91522.09
[pI]              9.32
[HOMOL]           TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA,
complete cds. 1e-166
[PROSITE]         MYRISTYL          6
[PROSITE]         CAMP_PHOSPHO_SITE      4
[PROSITE]         CK2_PHOSPHO_SITE      12
[PROSITE]         TYR_PHOSPHO_SITE       1
[PROSITE]         PKC_PHOSPHO_SITE      15
[PROSITE]         ASN_GLYCOSYLATION      7
[KW]              Alpha_Beta
[KW]              LOW COMPLEXITY      13.83 %
```

```

SEQ      MEEIKVLRRVKEENDRRGGFIRIFPTSETWEIYGSYLEHKTSMNYMLATRLFQDRGNPRR
SEG
PRD      ccchhhhhhhhhhhcccccceeecccccceeeccceeeccchhhhhhhhhhhccccc

```

```

SEQ      SLLTGRTRMTADGAPELKIESLNSKAKLHAALYERKLLSLEVRKRRRRSSRLRAMRPKYP
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccceccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

```

```

SEQ      VITQPAEMNVKTTETSEEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLTALVENT
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      ceecccchhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhcccccceeeccc

```

```

SEQ      PKNSMKVREWNNKGHCCKLETQELPKFNLMIQLQDNGNLSKMQARIAFSAYLQHVQI
SEG      .....
PRD      cccccceeeeeeccccccccchhhhhhhccchhhhhhhccchhhhhhhhhhhhhhhhhhhhh

```

```
SEQ      RLMKDSGGQTFSASWAAKEDQMELVVVFLKRASNNLQHSRLMVLPSRRLALLERRRIILA  
SEG      .....XXXXXXXXXXXXXXXXXXXX.  
PRD      hhhhccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhh
```

```

SEQ      HQLGDFIIYVKETEQAEMAEKSKKKKVEEEEEDGVNMENFQEFIRQASEAELEEVLTFYTG
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      hccccceeeehhhhhhhhhhhhhhhhhhhhhccccc hhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```
SEQ      KNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPPKQQQTTEIHSDKLSRF
SEG
PRD      cccccceccccccccccccccccccccccchhhhhhcccccccceeeeccccccc
```



```

SEQ  TTSAEKEAKLVYSNSSSGPTATLQKIPNTHLSSVTTSDLSPGPCHHSSLSQIPSAIPSPM
SEG  .....
PRD  hhhhhhhheeeecccccccccceeecccccccccccccccccccccccccccccccccccc

SEQ  HQPTILLNTVSASASPLHPGAQNIPTPTGLPCRSGSHTIGPFSSFQSAHIYSQKLSR
SEG  .....
PRD  cccccceccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhccc

SEQ  PSSAKAGSCYLNKHHSGIAKTQKEGEDASLYSKRYNQSMVTAEQLRLAEKQARQYSPSS
SEG  .....
PRD  cccccceeeecccccccccccccccccceeeecchhhhhhhhhhhhhhhhhhhhhhhccc

SEQ  HINLLTQQVTNLNLATGIINRSSASAPPTLRPIISPSGPTWSTQSDPQAPENHSSSPGSR
SEG  ..xxxxxxxxxxxxx.....
PRD  cccccccccccccccccccccccccccccceeecccccccccccccccccccccccccccc

SEQ  SLQTGGFAWEGEVENNVYSQATGVVPQHKYHPTAGSYQLQFALQQLQEQKLSRQLDQS
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD  cccccceeeecccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  RARHQAIIFSQTLPNSNLWTMNNAGCRISATASGQKPTTLQKVVPFPSSCASLVPKP
SEG  .....
PRD  hhhhhhhhhccccccccceeeccccceeeecccccccccceeeccccceeeccccc

SEQ  PPNHEQVLRRTSQAASNTRFRSSFNLYWYFFQAVS
SEG  .....
PRD  cccchhhhhhhhhhhccccccccccccceeeeeecccc

```

Prosites for DKFZphtes3_21n23.2

PS00001	221->225	ASN_GLYCOSYLATION	PDOC00001
PS00001	362->366	ASN_GLYCOSYLATION	PDOC00001
PS00001	381->385	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00001	620->624	ASN_GLYCOSYLATION	PDOC00001
PS00001	652->656	ASN_GLYCOSYLATION	PDOC00001
PS00004	106->110	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	271->275	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	789->793	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	109->112	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	185->188	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	543->546	PKC_PHOSPHO_SITE	PDOC00005
PS00005	561->564	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00005	629->632	PKC_PHOSPHO_SITE	PDOC00005
PS00005	793->796	PKC_PHOSPHO_SITE	PDOC00005
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	347->351	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	422->426	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	561->565	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00007	563->572	TYR_PHOSPHO_SITE	PDOC00007
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	248->254	MYRISTYL	PDOC00008
PS00008	510->516	MYRISTYL	PDOC00008
PS00008	557->563	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	756->762	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_21n23.2)

DKFZphtes3_22c23

group: testes derived

DKFZphtes3_22c23 encodes a novel 223 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, 3 EST hits (two from a testis library)

Sequenced by LMU

Locus: /map="9q34"

Insert length: 1113 bp

Poly A stretch at pos. 1073, polyadenylation signal at pos. 1055

```

1 GGTGGGCAAA GGCATCTTCC TCTGGGAAGG ACTGGCACAA GCACTTGGTC
51 CCTGGGGTTGT GTGCCTGGGA GGCCGGGATC AGGGCTGGCC CTCTTCTCC
101 CTGGCAAAGC AAAACCTCCC TTTTACTACT ATCAAGGGGA AGTAACTTGA
151 AGGTGCCTGT GGCAGGCAGC ACCTTGAGCC AACAGGAACC ATTGACATGC
201 GAGGCCCAGG GCAGGCAGAC TGTGCAGTGG CCATTGGGCG GCCCCTCGGG
251 GAGGTGGTGA CCTCCGCGT CCTTGAGAGT TCTCTCAACT GCAGTGCGGG
301 CGACATGTTG CTGCTTTGGG GCCGGCTCAC CTGGAGGAAG ATGTGCAGGA
351 AGCTGTTGGA CATGACTTTC AGCTCCAAGA CCAACACGCT GGTGGTGAGG
401 CAGCGCTGCG GCGGCCCAGG AGGTGGGGTG CTGCTGCGGT ATGGGAGCCA
451 GCTTGCTCCT GAAACCTTCT ACAGAGAATG TGACATGCAG CTCTTTGGGC
501 CCTGGGGTGA AATCGTGAGC CCCTCGCTGA GTCCAGCCAC GAGTAATGCA
551 GGGGGCTGCC GGTCTTTCAT TAATGTGGCT CCGCACGCAC GGATTGCCAT
601 CCATGCCCTG GCCACCAACA TGGGCGCTGG GACCGAGGGA GCCAATGCCA
651 GCTACATCTT GATCCGGGAC ACCCAGAGCT TGAGGACCAC AGCGTTCCAT
701 GGGCAGCAGG TGCTCTACTG GGAGTCAGAG AGCAGCCAGG CTGAGATGGA
751 GTTCAGCGAG GGCTTCCTGA AGGCTCAGGC CAGCCTGCGG GGCCAGTACT
801 GGACCCCTCCA ATCATGGGTA CCGGAGATGC AGGACCCCTCA GTCCTGGAAG
851 GGAAGGAAG GAACCTGAGG GTCATTGAAC ATTTGTTCCG TGTCTGGCCA
901 GCCCTGGAGG GTTGACCCCT GGTCTCAGTG CTTTCCAATT CGAACTTTT
951 CCAATCTTAG GTATCTACTT TAGAGTCTTC TCCAATGTCC AAAAGGCTAG
1001 GGGGTTGGAG GTGGGGACTC TGGAAAAGCA GCCCCCATTT CCTCGGGTAC
1051 CAATAAATAA AACATGCAGG CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA
1101 AAAAAAAAAA AAA

```

BLAST Results

Entry HSAC1644 from database EMBL:
 Genomic sequence from Human 9q34, complete sequence.
 Score = 2072, P = 8.8e-225, identities = 422/430
 5 exons Bp 41969-38232

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 197 bp to 865 bp; peptide length: 223
 Category: putative protein

```

1 MRPGQADCA VAIGRPLGEV VTLRVLESSL NCSAGDMLLL WGRLTWRKMC
51 RKLLDMTFSS KNTLVVRQR CGRPGGGVLL RYGSQAPET FYRECDMLF
101 GPWGEIVSPS LSPATSNAGG CRLFINVAPH ARIAIHALAT NMGAGTEGAN
151 ASYLIRDTLH SLRTTAFHGQ QVLYWESESS QAEMEFSEGF LKAQASLRGQ
201 YWTLQSWVPE MQDPQSWKKG EGT

```


No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22c23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_22c23, frame 2

Report for DKFZphtes3_22c23.2

```

[LENGTH]      223
[MW]           24546.19
[pI]           8.57
[PROSITE]      MYRISTYL      4
[PROSITE]      CK2_PHOSPHO_SITE      2
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           Alpha_Beta

SEQ      MRGPGQADCAVAIGRPLGEVTVLRVLESSLNCSAGDMLLLWGRLTWRKMCRKLLDMTFSS
PRD      cccccccccceeeccccceeeehhhhhccccccchhhhhhhchhhhhhhhhhhhhhhhhhhccc

SEQ      KTNLTLVVRQRCGRPGGVLLRYGSQLAPETFYRECDMLFGPWGEIVSPSLSPATSNAGG
PRD      cceeeeeeeccccccccceeeccccccchhhhhhhhhccccccccceeecccccccccccccc

SEQ      CRLFINVAPHARIAIHALATNMGAGTEGANASYILIRDTHSLRTAFHGGQVLYWESESS
PRD      ceeeeeeeccccceehhhhhhhhhccccccccceeeeeeccccceeeccccceeecccccc

SEQ      QAEMEFSEGFKAQSLRGQYWTLQSWVPQMOPQSWKGKEGT
PRD      hhhhhhhhhchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccc

```

Prosites for DKF2phtes3 22c23.2

PS00001	31->35	ASN_GLYCOSYLATION	PDOC00001
PS00001	150->154	ASN_GLYCOSYLATION	PDOC00001
PS00005	22->25	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	33->37	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	145->151	MYRISTYL	PDOC00008
PS00008	148->154	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3 22c23.2)

DKFZphtes3_22g2

group: nucleic acid management

DKFZphtes3_22g2 encodes a novel 1230 amino acid protein with nearly identical to rat TIP120.

TATA-binding protein TBP is a central component for transcriptional regulation and is a target for various transcription regulators. TBP-interacting protein 120 (TIP120) is a protein interacting with the TATA-binding protein (TBP). The novel protein is the human ortholog of rat TIP120. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP.

The new protein can find application in modulation of gene transcription.

KIAA0829, complete cds, nearly identical to rat TIP120

complete cDNA, complete cds, EST hits,

Sequenced by LMU

Locus: /map="387.3 cR from top of Chr12 linkage group"

Insert length: 5387 bp

Poly A stretch at pos. 5352, polyadenylation signal at pos. 5335

```

1 GGGAGCGAGT GCGGAGCGAG TGGGAGCGAG ACGGCCCTGA GTGGAAGTGT
51 CTGGCTCCCC GTAGAGGCCG TTCTGTACGC CCCGCCGCCG ATGAGCTCGT
101 TCTCACGCGA ACAGCGCCGT CGTTAGGCTG GCTCTGTAGC CTCGGCTTAC
151 CCCGGGACAG GCCCAGCCCT CGCCAGGGAG GGGGCGAGCC GTCGAGGCGC
201 CTCCTTAGTC AGCGTCGGCG TCGCGCTGCG ACCCTGGAAG CGGGAGCCGC
251 CGCGAGCGAG AGGAGGAGCT CCAGTGGCGG CGGCGGCGGC GGCAGCGGCA
301 GCGGGCAGCA GCTCCAGCAG CGCCAGCAGG CGGGATCGAG GCCGTCAACA
351 TGCGGAGCGC CTCGTACCAC ATTTCCAATT TGCTGGAAAA AATGACATCC
401 AGCGACAAGG ACTTTAGGTT TATGGCTACA AATGATTTGA TGACGGAAC
451 GCAGAAAGAT TCCATCAAGT TGGATGATGA TAGTGAAAGG AAAGTAGTGA
501 AAATGATTTT GAAGTTATTG GAAGATAAAA ATGGAGAGGT ACAGAATTTA
551 GCTGTCAAAAT GTCTGGTCC TTTAGTGAGT AAAGTGAAAG AATACCAAGT
601 AGAGACAATT GTAGATACCC TCTGCACTAA CATGCTTTCT GATAAAGAAC
651 AACTTTCGAGA CATTCAAGT ATTGGTCTTA AAACAGTAAT TGGAGAACTT
701 CCTCCAGCTT CCAGTGGCTC TGCATTAGCT GCTAATGTAT GTAAAAAGAT
751 TACTGGACGT CTTACAAGTG CAATAGCAAA ACAGGAAGAT GTCTCTGTT
801 AGCTAGAAGC CTTGGATATT ATGGCTGATA TGTGTAGCAG GCAAGGAGGA
851 CTTCTTGTTA ATTTCCATCC TTCAATTCTG ACCTGTCTAC TTCCCCAGTT
901 GACCAGCCCT AGACTTGCA TGAAGAAAAG AACCATTATC GCTCTTGGCC
951 ATCTGGTTAT GAGCTGTGGA AATATAGTTT TTGTAGATCT TATTGAACAT
1001 CTGTTGTGAG AGTTGTCCAA AAATGATTCT ATGTCAACAA CAAGAACCTA
1051 CATACAATGT ATTGCTGCTA TTAGTAGGCA AGCTGGTCAT AGAATAGGTG
1101 AATACCTTGA GAAGATAATT CCTTTGGTGG TAAATTTTGA CAATGTAGAT
1151 GATGATGAAT TAAGAGAGTA CTGTATTCAA GCCTTTGAAT CATTGTGAAG
1201 AAGATGTCTT AAGGAAGTAT ATCCTCATGT TTCTACCATT ATAAATATT
1251 GTCTTAAATA TCTTACCTAT GATCCAAATT ATAATTACGA TGATGAAGAT
1301 GAAGATGAAA ATGCAATGGA TGCTGATGGT GGTGATGATG ATGATCAAGG
1351 GAGTGATGAT GAATACAGTG ATGATGATGA CATGAGTTGG AAAGTGAGAC
1401 GTGCAGCTGC GAAGTGCTTG GATGCTGTAG TTAGCACAAG GCATGAAATG
1451 CTTCCAGAA TCTACAAGAC CCTCTCTCCT GCACTAATAT CCAGATTTAA
1501 AGAGCGTGAA GAGAAATGTA AGGCAGATGT TTTTCACGCA TACCTTTCTC
1551 TTTTGAAGCA AACTCGTCCT GTACAAAGT GGCTATGTGA CCGTGTGCA
1601 ATGGAGCAGG GAGAAACACC TTTAACAATG CTTCAGAGTC AGGTTCCCAA
1651 CATTGTTAAA GCTCTTCACA AACAGATGAA AGAAAAAGT GTGAAGACCC
1701 GACAGTGTG TTTTAAACATG TTAAGTGAGC TGGTAAATGT ATTACCTGGG
1751 GCCCTAACTC AACACATTCC TGTACTTGTA CCAGGAATCA TTTTCTCACT
1801 GAATGATAAA TCAAGCTCAT CGAATTTGAA GATCGATGCT TTGTCATGTC
1851 TATACGTAAT CCTCTGTAAC CATTCTCCTC AAGTCTTCCA TCCTCACGTT
1901 CAGGCTTTGG TTCCTCCAGT GGTGGCTTGT GTTGGAGACC CATTTTACAA
1951 AATTACATCT GAAGCACTTC TTGTTACTCA ACAGCTTGTC AAAGTAATTC
2001 GTCCCTTTAGA TCAGCCTTCC TCGTTTGATG CAACTCCTTA TATCAAAGAT
2051 CTATTTACCT GTACCATTAA GAGATTAAAA GCAGCTGACA TTGATCAGGA
2101 AGTCAAGGAA AGGGCTATTT CCTGTATGGG ACAAATTATT TGCAACCTTG
2151 GAGACAATTT GGGTCTGAC TTGCCTAATA CACTTCAGAT TTTCTGGAG
2201 AGACTAAAGA ATGAAATTAC CAGGTTAACT ACAGTAAAGG CATTGACACT
2251 GATTGCTGGG TCACCTTTGA AGATAGATTT GAGGCCCTGT CTGGGAGAAG
2301 GGGTTCCCTAT CCTTGCTTCA TTTCTTAGAA AAAACCAGAG AGCTTTGAAA
2351 CTGGGTACTC TTTCTGCCCT TGATATTCTA ATAAAAAAT ATAGTGACAG
2401 CTTGACAGCT GCCATGATTG ATGCAGTTCT AGATGAGCTC CCACCTCTTA
2451 TCAGCGAAGG TGATATGCAT GTTTCACAAA TGGCCATCAG TTTTCTTACC
2501 ACTTTGGCAA AAGTATATCC CTCCTCCCTT TCAAGATAAA GTGGATCCAT
2551 TCTCAATGAA CTTATTGGAC TTGTGAGATC ACCCTTATTG CAGGGGGGAG

```



```

2601 CTCTTAGTGC CATGCTAGAC TTTTCCAAG CTCTGGTGT CACTGGAACA
2651 AATAATTTAG GATACATGGA TTTGTTGCGC ATGCTGACTG GTCCAGTTTA
2701 CTCTCAGAGC ACAGCTCTTA CTCATAAGCA GTCTTATTAT TCCATTGCCA
2751 AATGTGTAGC TGCCCTTACT CGAGCATGCC CTAAGAGAGG ACCAGCTGTA
2801 GTAGGTCACT TTATTCAAGA TGTCAAGAAC TCAAGGTCTA CAGATTCCAT
2851 TCGTCTCTTA GCTCTACTTT CTCTTGAGA AGTTGGGCAT CATATTGACT
2901 TAAGTGGACA GTTGGAACTA AAATCTGTAA TACTAGAAGC TTTCTCATCT
2951 CCTAGTGAAG AAGTCAAATC AGCTGCATCC TATGCATTAG GCAGCATTAG
3001 TGTGGGCAAC CTTCTGAAT ATCTGCCGTT TGTCTTGCAA GAAATAACTA
3051 GTCAACCCAA AAGGCAGTAT CTTTACTTTC ATTCTTTGAA GGAAATTATT
3101 AGCTCTGCAT CAGTGGTGGG CCTTAAACCA TATGTTGAAA ACATCTGGGC
3151 CTTATTACTA AAGCACTGTG AGTGTGCAGA GGAAGGAACC AGAAATGTTG
3201 TTGCTGAATG TCTAGGAAAA CTAAGTCTAA TTGATCCAGA AACTCTCCTT
3251 CCACGGCTTA AGGGGTACTT GATATCAGGC TCATCATATG CCCGAAGCTC
3301 AGTGGTTACG GCTGTGAAAT TTACAATTTC TGACCATCCA CAACCTATTG
3351 ATCCACTGTT AAAGAACTGC ATAGGTGATT TCCTAAAAAC TTTGGAAGAC
3401 CCAGATTGTA ATGTGAGAAG AGTAGCCTTG GTCACATTTA ATTGAGCAGC
3451 ACATAACAAG CCATCATTA TAAGGGATCT ATTGGTACT GTTCTTCCAC
3501 ATCTTTACAA TGAACAAAAA GTTAGAAAGG AGCTTATAAG AGAGGTAGAA
3551 ATGGGTCCAT TTAACATAC GGTGATGAT GGTCTGGATA TTAGAAAGGC
3601 AGCATTGAG TGTATGTACA CACTTCTAGA CAGTTGTCTT GATAGACTTG
3651 ATATCTTTGA ATTTCTAAAT CATGTTGAAG ATGGTTTGAA GGACCATTAT
3701 GATATTAAAG TGCTGACATT TTTAATGTTG GTGAGACTGT CTACCCTTGT
3751 TCCAAGTGCA GTACTGCAGA GGTGGACCG ACTTGTGAG CCATTACGTG
3801 CAACATGTAC AACTAAGGTA AAGGCAAACT CAGTAAAGCA GGAGTTTGAA
3851 AAACAAGATG AATTAAAGCG ATCTGCCATG AGAGCAGTAG CAGCACTGCT
3901 AACCATTCCA GAAGCAGAGA AGAGTCCACT GATGAGTGAA TTCCAGTCAC
3951 AGATCAGTTC TAACCCCTGAG CTGGCGGCTA TCTTTGAAAG TATCCAGAAA
4001 GATTTCATCAT CTACTAACTT GGAATCAATG GACACTAGTT AGATGTTTGT
4051 TCACCATGGG GACCATTACA TATGACCATA CAATGCACGT AATTGACAGG
4101 TTAATCATAA GACATGGAAA GAGAAGTGTC TAAAAGCTTC AAAATGTTCC
4151 ACTTTTTTTT CCTTCATGGA GACTGTTTGT TTGGCTTTCT TCCATTGTTG
4201 TTTTGTAGC ATTTATTTC GAAATGTGTA TTTCCATAAT CCAGAGGTTG
4251 TAAAACCACT AGTGTTTAG TGGTTACAGC AACATTTGAA ATGGAAACTA
4301 AAAGTTAGGA TTTTATGGAG TATGGAGATA GGGTCCAGTA TCTATTACC
4351 CTGTAATGTT TAGGATTAAA ATGTTAAAT TTTGTGACCA TGAATTTCTT
4401 TCTTTTATAA ATTTCTCAT TTAATAATCA AAAATCTTGC AAAACAAAAA
4451 CCATGTTTCT TTTTCTTGT TAACTTTTTG TTTTCAGCAA CATAAATTGA
4501 TTTTGTAGCT GCAGACAGA ATATCCATAT AAGATTGTT AACCATTTC
4551 GAGAGTTTGG CAATTTTAA AAGATAATAA GGTATCATTT TTAAGTATGA
4601 AAATTAACAA TATCCCTGTT GCGCACATA ATTTGTCATG AGTAAGTTTA
4651 CAAATATGTA TCGTCTGTA AGCAGCATGT GCAGATTATT CATAATATAG
4701 AAGTTAAAT AAGTATTAGT GCAATTTTCA GATATTTATT TTGTCACAGA
4751 AAACACATTA TCTGGAGAGA AAGAAAGGAG AATTTTGTAG ACTTGGGTTT
4801 TCTTAATGCC AGTGTGAATT TGCAGATGTT TTCAGAAAT CAAGTCACAG
4851 TAACAATTTG CCACTTTTTT CTATTATAAA TCTTCTTACT TAAATTTTGA
4901 ATATTTAGTT TTTCTCAGTT ACCCATTTGT GTGTGTGTGA TTCCACTTAG
4951 AAATTCCTAA AACCAGATTT TTCTTTCATT CCGTTTGGAT GTCTACATT
5001 CTTATCAAAG GATATAAATA CTGTGTATGC TTTTGAATTT TATTTTTAGG
5051 AAAATTCGTA AGCCAGCTAT CACAGGTTTG TTAGCTAATA ATAGTATTTT
5101 CTTTATGTTG AGTTAGGTTT TTCCCATCT CCGTGTAGAGC GAATTTACAT
5151 ATTGTATTGG GTAAGTGTTT ACTACTTTTC CTGATTAGG GATCTGTGCT
5201 GGGGAACAAA GCTTTTGAGC TACCTTATAT TGTAGTTAAA ATTTTATTTA
5251 ACATATCCTT CAGTGAGCTC ATTTCACT GTAGCCTCTT CCTTAAATTT
5301 TGTGGTGCTC CTGTAACAGT AAGAACTAAT TCTGAAATAA AAGACATCTC
5351 CTAAAAAAA AAAAAAAA AAAAAAAA AAAAAA

```

BLAST Results

Entry HS793345 from database EMBL:
human STS WI-12457.
Score = 1985, P = 1.3e-83, identities = 433/460

Medline entries

97127450:
Molecular cloning of a novel 120-kDa TBP-interacting
protein.

Peptide information for frame 2

ORF from 350 bp to 4039 bp; peptide length: 1230

Category: known protein
 Classification: Nucleic acid management

```

1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV
51 KMILKLEDK NGEVQNLAVK CLGPLVSKVK EYQVETIVDT LCTNMLSDKE
101 QLRDISSIGL KTVIGELPPA SSGSALAANV CKKITGRITS AIAKQEDVSV
151 QLEALDIMAD MLRSRQGLLV NFHPSILTCL LPQLTSPRLA VRKRTIIALG
201 HLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA ISRQAGHRIG
251 EYLEKIIPLV VKFCNVDDDE LREYCIQAFE SFVRRCPKEV YPHVSTIINI
301 CLKYLTYPN YNYDDEDEDE NAMDADGGDD DDQGSDDDEYS DDDDMSWKVR
351 RAAAKCLDAV VSTRHEMLPE FYKTVSPALI SRFKEREENV KADVHFAYLS
401 LLKQTRPVQS WLCDDPAMEQ GETPLTMLQS QVPNIVKALH KQMEKSVKT
451 RQCCFNMLTE LVNVLPALT QHIPVLVPGI IFSLNKSSS SNLKIDALSC
501 LYVILCNHSP QVFHPPHVAL VPPVAVACVD PFYKITSEAL LVTQQLVKVI
551 RPLDQPSFDF ATPYIKDLFT CTIKRLKAAD IDQEVKERAI SCMGQIICNL
601 GDNIGSDLPN TLQIFLERLK NEITRLTTVK ALTLIAGSPL KIDLRPVLGE
651 GVPILASFLR KNQRAKLGK LSALDILIKN YSDSLTAAMI DAVLDELPLP
701 ISEDMHVQSQ MAISELTTLA KVPSSLSKI SGSIINELIG LVRSPLLQGG
751 ALSAMLOFFQ ALVVTGTNNL GYMDLLRMLT GPVYSQSTAL THKQSYYSIA
801 KCVAAALTRAC PKEGPAVVGQ FIQDVKNRSR TDSIRLLALL SLGEVGHMID
851 LSGQLELKSV ILEAFSSPSE EVKSAASYAL GSISVGNLPE YLPFVLQEI
901 SQPKRQYLLL HSLKEIISA SVVGLKPYVE NIWALLLKH CEAEGTRNV
951 VAECLGKLT IDPETLLPRL KGYLISGSSY ARSSVVTAVK FTISDHPQPI
1001 DPLLKNCIGD FLKTLEDDPDL NVRRVALVTF NSAAHNKPSL IRDLLDTVLP
1051 HLYNETKVRK ELIREVEMGP FKHTVDDGLD IRKAAFECEM TLLDSCDLRL
1101 DIFEFLNHVE DGLKDHYDIK MLTFLMLVRL STLCPSAVLQ RLDRLVEPLR
1151 ATCTTKVKAN SVKQEFKQD ELKRSAMRAV AALLTIPEAE KSPLMSEFQS
1201 QISSNPELAA IFESIQKDS STNLESMDTS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22g2, frame 2

TREMBL:AB020636_1 gene: "KIAA0829"; product: "KIAA0829 protein"; Homo sapiens mRNA for KIAA0829 protein, partial cds., N = 1, Score = 5986, P = 0

TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds., N = 1, Score = 6203, P = 0

>TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds.
 Length = 1,230

HSPs:

Score = 6203 (930.7 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 1227/1230 (99%), Positives = 1228/1230 (99%)

```

Query:      1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDK 60
            MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDK 60
Sbjct:      1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDK 60

Query:      61 NGEVQNLAVK CLGPLVSKVKEYQVETIVDT LCTNMLSDKEQLRDISSIGL KTVIGELPPA 120
            NGEVQNLAVK CLGPLVSKVKEYQVETIVDT LCTNMLSDKEQLRDISSIGL KTVIGELPPA 120
Sbjct:      61 NGEVQNLAVK CLGPLVSKVKEYQVETIVDT LCTNMLSDKEQLRDISSIGL KTVIGELPPA 120

Query:      121 SSGSALAANV CKKITGRITS AIAKQEDVSVQLEALDIMAD MLRSRQGLLVNFHPSILTCL 180
            SSGSALAANV CKKITGRITS AIAKQEDVSVQLEALDIMAD MLRSRQGLLVNFHPSILTCL 180
Sbjct:      121 SSGSALAANV CKKITGRITS AIAKQEDVSVQLEALDIMAD MLRSRQGLLVNFHPSILTCL 180

Query:      181 LPQLTSPRLAVRKRRTI IALGHLVMSCGNIVFVDLIEHLL SELSKNDSMSTTRTYIQCIAA 240
            LPQLTSPRLAVRKRRTI IALGHLVMSCGNIVFVDLIEHLL SELSKNDSMSTTRTYIQCIAA 240
Sbjct:      181 LPQLTSPRLAVRKRRTI IALGHLVMSCGNIVFVDLIEHLL SELSKNDSMSTTRTYIQCIAA 240

Query:      241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI 300
            ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI 300
Sbjct:      241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI 300

Query:      301 CLKYLTYPN YNYDDEDEDE NAMDADGGDDDDQGSDDDEYS DDDDMSWKVRRAAAKCLDAV 360
            CLKYLTYPN YNYDDEDEDE NAMDADGGDDDDQGSDDDEYS DDDDMSWKVRRAAAKCLDAV 360
Sbjct:      301 CLKYLTYPN YNYDDEDEDE NAMDADGGDDDDQGSDDDEYS DDDDMSWKVRRAAAKCLDAV 360

Query:      361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWLCPDAMEQ 420
            VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWLCPDAMEQ 420
Sbjct:      361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWLCPDAMEQ 420

```


Query: 421 GETPLTMLQSQVPNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLVPGI 480
 GETPLTMLQSQVPNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLVPGI
 Sbjct: 421 GETPLTMLQSQVPNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLVPGI 480

Query: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVAVCVGDPFYKITSEAL 540
 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVAVCVGDPFYKITSEAL
 Sbjct: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVAVCVGDPFYKITSEAL 540

Query: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600
 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
 Sbjct: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600

Query: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR 660
 GDNLG DL NTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR
 Sbjct: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR 660

Query: 661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPSESMDHVSQMAISFLTTLA 720
 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPSESMDHVSQMAISFLTTLA
 Sbjct: 661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPSESMDHVSQMAISFLTTLA 720

Query: 721 KVPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT 780
 KVPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT
 Sbjct: 721 KVPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT 780

Query: 781 GPVYSQSTALTHKQSYYSIAKVAALTRACPKEGPAVVGQFIQDVKNRSTDSIRLLALL 840
 GPVYSQSTALTHKQSYYSIAKVAALTRACPKEGPAVVGQFIQDVKNRSTDSIRLLALL
 Sbjct: 781 GPVYSQSTALTHKQSYYSIAKVAALTRACPKEGPAVVGQFIQDVKNRSTDSIRLLALL 840

Query: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900
 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
 Sbjct: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900

Query: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIWALLLKHCECAEEGTRNVVAECLGKLT 960
 SQPKRQYLLHSLKEIISASVVGKPYVENIWALLLKHCECAEEGTRNVVAECLGKLT
 Sbjct: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIWALLLKHCECAEEGTRNVVAECLGKLT 960

Query: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL 1020
 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL
 Sbjct: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL 1020

Query: 1021 NVRRVALVTFNSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080
 NVRRVALVTFNSAAHNKPSLIRDLLD+VLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
 Sbjct: 1021 NVRRVALVTFNSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080

Query: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140
 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
 Sbjct: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140

Query: 1141 RLDRLEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200
 RLDRLEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
 Sbjct: 1141 RLDRLEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200

Query: 1201 QISSNPELAAIFESI QKDSSTNLESMDTS 1230
 QISSNPELAAIFESI QKDSSTNLESMDTS
 Sbjct: 1201 QISSNPELAAIFESI QKDSSTNLESMDTS 1230

Pedant information for DKF2phtes3_22g2, frame 2

Report for DKF2phtes3_22g2.2

[LENGTH] 1230
 [MW] 136376.58
 [pI] 5.52
 [HOMOL] TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for
 TIP120, complete cds. 0.0
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 5.28 %

SEQ MASASYHISNLEKMTSSDKDFRMTNDLMTLQKDSIKLDDSERKVVKMILKLEDD
 SEG
 PRD cccccchhhhhhhhhccccceeeehhhhhhhhhccccccccchhhhhhhhhhhcc
 MEM

SEQ NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTIVIGELPPA
 SEGxxxx
 PRD cccccceeeeeeceeeehhhhhhhhhhhccchhhhhccccccccchhhhhhhhhccccc

MEM
SEQ SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL
SEG xxxxxxxx.....
PRD cccccchhhhhhhccchhhhhhhccccchhhhhhhhhhhhhhhccceeeccchhhhh
MEM
SEQ LPQLTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDMSMSTTRTYIQCIAA
SEG
PRD hccccchhhhhhhhhhhheeeccceehhhhhhhhhhhcccccchhhhhhhhhhh
MEMMMMMMMMMMMMMMMMM.....
SEQ ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPEVYPHVSTIINI
SEG
PRD hhhhhccccccccchhhhhhhheeeccchhhhhhhhhhhhhhhhhccccceecchhhhh
MEM
SEQ CLKYLTYPNINYDDEDEDENAMADGGDDDDQGSDDDEYSDDDMSWKVRRAAKCLDAV
SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD hhhhccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhh
MEM
SEQ VSTRHEMLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCDPDAMEQ
SEG
PRD hhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeecccccc
MEM
SEQ GETPLTMLQSQVPNIVKALHKQMKESVKTRQCCFNMLTELNVNLPGALTQHIPVLVPGI
SEG
PRD cccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccceeeccce
MEM
SEQ IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPHVQALVPPVACVGDPPFYKITSEAL
SEGXXXXXXXXXXXXXXXXXXXX.....
PRD eeeccccccccchhhhhhhheeeccccccccccceeeccceeeccchhhhhhhhh
MEM
SEQ LVTQQLVKVIRPLDQPSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
SEG
PRD hhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhheeecc
MEM
SEQ GDNLGSDLPTNLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPIIASFLR
SEG
PRD cccccccccchhhhhhhhhccchhhhhhhhhhhheeeccccccccceehhhhhhhhhhh
MEM
SEQ KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLISESDMHVSQMAISFLTTLA
SEG
PRD hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhh
MEM
SEQ KVYPSSLSKISGISILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT
SEG
PRD cccccceecchhhhhhhhhhhccccccccchhhhhhhhhhhheeeccccchhhhhhhhhc
MEM
SEQ GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVQGFIQDVKNRSRSTDSIRLLALL
SEG
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccchhhhhhhhh
MEM
SEQ SLGEVGHIDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
SEG
PRD hccccccccccccccccceeecccccchhhhhhhhhhhccccccccccccchhhhhhhhh
MEM
SEQ SQPKRQYLLHSLKEIISASVVGKPYVENIWALLLKHCECAEEGTRNVVAECLGKLT
SEG
PRD cccchhhhhhhhhhhhhhhccccceehhhhhhhhhhhhhhhccccceeecccccccc
MEM
SEQ IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD
SEG
PRD cccccccccccccccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhcccc
MEM
SEQ NVRRVALVTFNSAAHNKPSLIRDLLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
SEG
PRD cccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc
MEM


```
SEQ  IRKAAFECCMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
SEG  .....
PRD  hhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RLDRLVEPLRATCTTKVKANSVKQEFQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
SEG  .....
PRD  hhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  QISSNPELAAIFESIQKDSSTNLESMDTS
SEG  .....
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....
```

(No Prosite data available for DKFZphtes3_22g2.2)

(No Pfam data available for DKFZphtes3_22g2.2)

DKFZphtes3_22n13

group: testes derived

DKFZphtes3_22n13 encodes a novel 677 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

dJ1042K10.3, complete

Sequenced by LMU

Locus: /map="22q13.1-13.2"

Insert length: 3353 bp

Poly A stretch at pos. 3315, polyadenylation signal at pos. 3298

```
1 ATGGAACCCAC TATCCCCACT GCCAAGTCCA CCCCCACACT CATTAAAGCAA
51 AGCCAACCCA AGTCTGCCAG TGAGAAGTCA CAGCGCAGCA AGAAGGCCAA
101 GGAGCTGAAG CCAAAGGTGA AGAAGCTCAA GTACCACCAG TACATCCCCC
151 CGGACCAGAA GCAGGACAGG GGGGCACCCC CCATGGACTC ATCCTACGCC
201 AAGATCCTGC AGCAGCAGCA GCTCTTCCTC CAGCTGCAGA TCCTCAACCA
251 GCAGCAGCAG CAGCACCACA ACTACCAGGC CATCCTGCCT GCCCCGCCAA
301 AGTCAGCAGG CGAGGCCCTG GGAAGCAGCG GGACCCCCC AGTACGCAGC
351 CTCTCCACTA CCAATAGCAG CTCCAGCTCG GCGCCCCCTG GGCCCTGTGG
401 GCTGGCAGCT CAGAACAGCA CCTCACTGAC TGGCAAGCCG GGAGCCCTGC
451 CGGCCAACCT GGACGACATG AAGGTGGCAG AGCTGAAGCA GGAGCTGAAG
501 TTGCGATCAC TGCCTGTCTC GGGCACCAAA ACTGAGCTGA TTGAGCGCCT
551 TCGAGCCTAT CAAGACCAA TCAGCCCTGT GCCAGGAGCC CCCAAGGCCC
601 CTGCCGCCAC CTCTATCCTG CACAAGGCTG GCGAGGTGGT GGTAGCCTTC
651 CCAGCGGCCC GGCTGAGCAC GGGGCCAGCC CTGGTGGCAG CAGGCCCTGGC
701 TCCAGCTGAG GTGGTGGTGG CCACGGTGGC CAGCAGTGGG GTGGTGAAGT
751 TTGGCAGCAG GGGCTCCACG CCCCCCGTGT CTCCCACCCC CTCGGAGCGC
801 TCACTGCTCA GCACGGGCGA TGAAAATCC ACCCCCGGGG ACACCTTTGG
851 TGAGATGGTG ACATCACCTC TGACGCAGCT GACCTGCGAG GCCTCGCCAC
901 TGCAGATCCT CGTGAAGGAG GAGGGCCCCC GGGCCGGGTC CTGTTGCCTG
951 AGCCCTGGGG GGGCGGCGGA GCTAGAGGGG CGCGACAAGG ACCAGATGCT
1001 GCAGGAGAAA GACAAGCAGA TCGAGGCGCT GACGCGCATG CTCGGGCAGA
1051 AGCAGCAGCT GGTGGAGCGG CTCAGCTGCG AGCTGGAGCA GGAGAAGCGA
1101 GCCCAGCAGC CCGCCCCCGC CCCCAGCCCC CTCGGCACCC CCGTGAAGCA
1151 GGAGAACAGC TTCTCCAGCT GCCAGCTGAG CCAGCAGCCC CTGGGCCCGG
1201 CTCACCCATT CAACCCAGC CTGGCGGCCC CAGCCACCAA CCACATAGAC
1251 CCTTGTGCTG TGGCCCCAGG GCCCCCGTCC GTGGTGGTGA AGCAGGAAGC
1301 CTTGCAGCCT GAGCCCCGAG CGGTCCCCGC CCCCAGTTG CTTCTGGGGC
1351 CTCAGGGCCC CGGCCCTCATC AAGGGGGTTG CACCTCCCAC CCTCATCACC
1401 GACTCCACAG GGACCCACCT TGTCCTCACC GTGACCAATA AGAATGCAGA
1451 CAGCCCTGGC CTGTCCAGTG GGAGCCCCCA GCAGCCCTCG TCCCAGCCTG
1501 GCTCTCCAGC GCCTGCCCCC TCTGCCCAGA TGGACCTGGA GCACCCACTG
1551 CAGCCCCCTCT TTGGGACCCC CACTTCTCTG CTGAAGAAGG AACCACCTGG
1601 CTATGAGGAA GCCATGAGCC AGCAGCCCCA ACAGCAGGAA AATGGTTCTT
1651 CAAGCCAGCA GATGGACGAC CTGTTTGACA TTCTCATTCA GAGCGGAGAA
1701 ATTTTCAGCAG ATTTCAAGGA GCCGCCATCC CTGCCAGGGA AGGAGAAGCC
1751 ATCCCCGAAG ACAGTCTGTG GGTCCCCCCT GGCAGCACAG CCATCACCTT
1801 CTGCTGAGCT CCCCAGGCT GCCCCACCTC CTCCAGGCTC ACCCTCCCTC
1851 CCTGGAGGCC TGGAGGACTT CCTGGAGAGC AGCACGGGGC TGCCCTGCTT
1901 GACCAAGTGG CATGACGGGC CAGAGCCCCC TTCCCTCATT GACGACCTCC
1951 ATAGCCAGAT GCTGAGCAGC ACTGCCATCC TGGACCACCC CCGCTCACCC
2001 ATGGACACCT CGGAATTGCA CTTTGTTCCT GAGCCCAGCA GCACCATGGG
2051 CCTGGACCTG GCTGATGGCC ACCTGGACAG CATGGACTGG CTGGAGCTGT
2101 CGTCAGGTGG TCCCGTGCTG AGCCTAGCCC CCCTCAGCAC CACAGCCCCC
2151 AGCCTCTTCT CCACAGACTT CCTCGATGGC CATGATTTCG AGCTGCACTG
2201 GGATTCTCTG TTGTAGCTCT CTGGCTCAAG ACGGGGTGGG GAAGGGGCTG
2251 GGAGCCAGGG TACTCCAATG CGTGGCTCTC CTGCGTGATT CGGCCCTCTC
2301 ACATGGTTGT GAGTCTTGAC AATCACAGCC CTGCTTTTTT CCCTTCCCTG
2351 GGAGGCTAGA ACAGAGAAGC CCTTACTCCT GGTTCAAGTC CACGCAGGGC
2401 AGAGGAGAGC AGCTGTCAAG AAGCAGCCCT GGCTCTCAGC CTGGGGTTTT
2451 GGACACACGG TCAGGGTCAG GGCCATTTCG GCTTGACCTC CTTTTTTGAG
2501 GTCAGGGGGC ACTGTCTGTC TGGCTACAAT TTGGCTAAGG TAGGTGAAGC
2551 CTGGCCAGGC GGGAGGCTTC TCTTCTGACC CAGGGCTGAG ACAGGTTAAG
2601 GGGTGAATCT CCTTCCTTTC TCTCCCTGCT TTGCTGTGAA GGGAGAAATT
2651 AGCCTGGGCC TCTACCCCTT ATCCCTGTG TCTGCCAACC CCAGGATCCC
2701 AGGGCTCCCT GCCATTTTAG TGTCTTGGTG TAGTGTAAAC ATTTAGTGGT
2751 TGGTGGCAAC AATTTTATGT ACAGGTGTAT ATACCTCTAT ATTATATATC
2801 GACATACATA TATATTTTTG GGGGGGGGCG GACAGGAGAT GGGTGCAACT
```



```

2851 CCCTCCCATC CTACTCTCAC AGAAGGGCCT GGATGCAAGG TTACCCCTTGA
2901 GCTGTGTGCC ACAGTCTGGT GCCCAGTCTG GCATGCAGCT ACCCAGGCCC
2951 ACCCATCACG TGTGATTGAC ATGTAGGTAC CCTGCCACGG CCTATGCCCC
3001 ACCTGCCCTG CTTCTGGCT CCTTATCAGT GCCATGAGGG CAGAGGTGCT
3051 ACCTGGCCTT CTGCCCAGGA GCTCTCCACC CACTCACATT CCGTCCCCGC
3101 CGCCTCACTG CAGCCAGCGT GGCCTAGGA CAGGAGGAGC TTCGGGCCCCA
3151 GCTTCACCTT GCGGTGGGGC TGAGGGGTGG CCATCTCCTG CCCTGGGGCC
3201 ACTGGCTTCA CATTCTGGGC TGACTCATAG GGGAGTAGGG GTGGAGTCAC
3251 CAAAACCACT GCTGGGACAA AGATGGGGAA GGTGTGTGAA CTTTTTAAAA
3301 TAAACACAAA AACACAGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAG

```

BLAST Results

Entry HS1042K10 from database EMBL:

Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2.

Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island.

Score = 7997, P = 0.0e+00, identities = 1617/1645

7 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 183 bp to 2213 bp; peptide length: 677

Category: similarity to unknown protein

Classification: unclassified

```

1 MDSSYAKILQ QQQLFLQLQI LNQQQQQHHN YQAILPAPPK SAGEALGSSG
51 TPPVRSLSST NSSSSSGAPG PCGLARQNST SLTGKPGALP ANLDDMKVAE
101 LKQELKLRLS PVSQTKTELI ERLRAYQDOI SPVPGAPKAP AATSILHKAG
151 EVVVAFFPAAR LSTGPALVAA GLAPAEVVVA TVASSGVVKF GSTGSTPPVS
201 PTPSERSLLS TGDENSTPGD TFGEMVTSPL TQLTLQASPL QILVKEEGPR
251 AGSCCLSPGG RAELEGRDKD QMLQEKDKQI EALTRMLRQK QQLVERLKLQ
301 LEQEKRAQQP APAPAPLGTP VKQENSFSSC QLSQQPLGPA HPFNPSLAAP
351 ATNHIDPCAV APGPPSVVVK QEALQPEPEP VPAPQLLLGP QGPGLIKGVA
401 PPTLITDSTG THLVLTVTNK NADSPGLSSG SPQQPSSQPG SPAPAPSAQM
451 DLEHFLQLPL GTPTSLLKKE PPGYEEAMSQ QPKQENGSS SQQMDDLFDI
501 LIQSGEISAD FKEPPSLPGK EKPSPKTVCG SPLAAQSPSP AELPQAAPPP
551 PGSPSLPGRG EDFLESSTGL PLLTSGHDGP EPLSLIDDLH SQMLSSTAIL
601 DHPSPMDTS ELHFVPEPSS TMGLDLADGH LDSMDWLELS SGGPVLSLAP
651 LSTTAPSLFS TDFLDGHDQ LHWDSCL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22n13, frame 3

TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island., N = 1, Score = 1285, P = 4.9e-131

TREMBL:CEUK06A9_3 gene: "K06A9.1a"; Caenorhabditis elegans cosmid K06A9., N = 2, Score = 149, P = 1.3e-09

TREMBLNEW:SSI132828_1 product: "p210 protein"; Spermatozopsis similis mRNA for p210 protein, partial, N = 1, Score = 171, P = 2.8e-09

>TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC

Length = 243

HSPs :

Score = 1285 (192.8 bits), Expect = 4.9e-131, P = 4.9e-131
Identities = 243/243 (100%), Positives = 243/243 (100%)

Query:	435	PSSQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLKKEPPGYEEAMSQPKQKQENGSSSQM	494
		PSSQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLKKEPPGYEEAMSQPKQKQENGSSSQM	
Sbjct:	1	PSSQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLKKEPPGYEEAMSQPKQKQENGSSSQM	60
Query:	495	DDLFDILIQSGEISADFKPEPSLPGKEKPSPKTVCGSPLAAQSPSPAELPQAAPPPPGSP	554
		DDLFDILIQSGEISADFKPEPSLPGKEKPSPKTVCGSPLAAQSPSPAELPQAAPPPPGSP	
Sbjct:	61	DDLFDILIQSGEISADFKPEPSLPGKEKPSPKTVCGSPLAAQSPSPAELPQAAPPPPGSP	120
Query:	555	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQLMSSTAILDHPSPMDTSELHF	614
		SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQLMSSTAILDHPSPMDTSELHF	
Sbjct:	121	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQLMSSTAILDHPSPMDTSELHF	180
Query:	615	VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	674
		VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	
Sbjct:	181	VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	240
Query:	675	SCL 677	
		SCL	
Sbjct:	241	SCL 243	

Pedant information for DKFZphtes3_22n13, frame 3

Report for DKFZphtes3 22n13.3

```
[LENGTH]          677
[MW]               70743.01
[pI]               4.93
[HOMOL]            TREMBL:HS1042K10.6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)";
Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for
Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with
probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative
CpG island. 1e-111
[KW]               TRANSMEMBRANE 1
[KW]               LOW COMPLEXITY 21.57 %
[KW]               COILED COIL 4.58 %
```

```

SEQ      MDSSYAKILQQOQLFLQLQILNQOQQOHHNYQAILPAPPKSAGEALGSSGTPPVRSLSTT
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx ..... xxxxxx
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhccceeeeeeccccceeeccccccccceeeccc
COILS
MEM      .....

SEQ      NSSSSSGAPGPCGLARQNSTSLTGKPGALPANLDDMKVAELKQELKLRSLPVSGTKTELI
SEG      xxxxxxxx ..... xxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccceccccccccccccccccccccchhhhhhhhhhhhhhhcccccchhhh
COILS
MEM      .....

SEQ      ERLRAYQDQISPVPGAPKAPAATSILHKAGEVVVAFPAARLSTGPALVAAGLAPAEVVVA
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhccccccccccccceeeeeeeccceeeccccccccccccccccceeeeee
COILS
MEM      ..... MMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      TVASSGVVKFGSTGSTPPVSPTPSERSLLSTGDENSTPGDTPGEMVTSPLTQLTLQASPL
SEG      xxxxxxxx . xxxxxxxxxxxxxxxxxxxx .....
PRD      eeeccccccccccccccccccccccccceeeccccccccccccccccceeeccccceeecccce
COILS
MEM      M.....

SEQ      QILVKEEGPRAGSCCLSPGGRAELEGRDKDOMLOEKDKQIEALTRMLRQKQQLVERLKLO
SEG      .....
PRD      eeeeeccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS
MEM      ..... cccccccccccccccccccccccc

SEQ      LEQEKRAQQPAPAPAPLGTVPVKQENSFSSCQLSQPLGPAHPFNPSLAAPATNHIDPCA

```



```
SEG      .....xxxxxxxxx.....
PRD      hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    CCCCCC.....
MEM      .....

SEQ      APGPPSVVVKQALQPEPEPVPAQQLLGPGGLIKGVAPPTLITDSTGTHLVLTVTNK
SEG      .....xxxxxxxxx.....
PRD      cccccceeecccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      NADSPGLSSGSPQQPSSQPGSPAPAPSAQMDLEHPLQLFGTPTSLLKKEPPGYEEAMSQ
SEG      ...xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      QPKQQENGSSSQMDDLFDILIQSGEISADFEKPPSLPGKEKPSPKTVCGSPLAAQSPSPS
SEG      .....xxxxxxxxx.....
PRD      cccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      AELPQAAPPPPGSPSLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAIL
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhccceee
COILS    .....
MEM      .....

SEQ      DHPPSPMDTSELHFVPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFS
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccceeeccccceeecccccccccc
COILS    .....
MEM      .....

SEQ      TDFLDGHDQLQHWDSCL
SEG      .....
PRD      ccccccccccecccccc
COILS    .....
MEM      .....
```

(No Prosite data available for DKFZphtes3_22n13.3)

(No Pfam data available for DKFZphtes3_22n13.3)

DKFZphtes3_23111

group: intracellular transport and trafficking

DKFZphtes3_23111 encodes a novel 186 amino acid protein nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).

Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors (ARFs) to donor membranes, leading to recruitment of cocatomer, bud formation, and eventual vesicle release. ARFs are approximately 20-kDa GTPases that are active with bound GTP and inactive with GDP bound. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and seems to be a novel ARF. It seems to have an important role in vesicular transport and vesicular trafficking.

The new protein can find application in modulating vesicle transport and trafficking in cells.

nearly identical to mouse Arl6, ADP-ribosylation-like factor homolog

start at Bp 15 matches kozak consensus ANNatgG

Sequenced by LMU

Locus: unknown

Insert length: 717 bp

Poly A stretch at pos. 689, no polyadenylation signal found

```

1 ATTTGAATCA CATTATGGGA TTGCTAGACA GACTTTCAGT CTTGCTTGGC
51 CTGAAGAAGA AGGAGGTTCA TGTTTGTGTC CTTGGGCTAG ATAATAGTGG
101 CAAAACGACG ATCATTAAAC AACTTAAACC TTCAAATGCT CAATCTCAAA
151 ATATCCTTCC AACAAATAGGA TTCAGCATAG AGAAATTCAA ATCATCCAGT
201 TTGTCAATTA CAGTGTGTTGA CATGTCAGGT CAAGGAAGAT ACAGAAATCT
251 CTGGGAACAC TATTATAAAG AAGGCCAAGC TATTATTTT GTCTTGTGATA
301 GTAGTGATAG ATTAAGAATG GTTGTGGCCA AAGAAGAACT CGATACTCTT
351 CTGAATCATC CAGATATTAA ACACCGTCGA ATTCCAATCT TATTCTTTGC
401 AAATAAAATG GATCTTAGAG ATGCAGTGAC ATCTGTAAAA GTGTCTCAGT
451 TGCTGTGTTT AGAGAACATC AAAGATAAAC CCTGGCATAT TTGTGCTAGT
501 GATGCCATAA AAGGAGAAGG CTGCAAGAA GGTGTAGACT GGCTTCAAGA
551 TCAGATCCAG ACTGTGAAGA CATGAAAAGA TAATAGTTGG AAACCTCAGC
601 AATTTTCAAT TCAAGGAATC TATCTAAGAC AAATAGAATA CATTTTGTAA
651 AAGATGTTTA TGCATCAAAA AATATAATTT TCTGCTTGCA AAAAAAAAAA
701 AAAAAAAAAA AAAAAAG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 15 bp to 572 bp; peptide length: 186
 Category: strong similarity to known protein
 Classification: Intracellular transport and traffic
 Prosite motifs: ATP_GTP_A (24-32)

```

1 MGLLDRLSVL LGLKKKEVHV LCLGLDNSGK TTIINKLKPS NAQSQNILPT
51 IGFSIEKFKS SLSFTVFDM SQGGRYRNW EHYYKEQAI IFVIDSSDRL
101 RMVVAKEELD TLLNHPDIKH RRIPILFFAN KMDLRDAVTS VKVSQLLCLE
151 NIKDKPWHIC ASDAIKGEGL QEGVDWLQDQ IQTVKT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23111, frame 3

TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds., N = 1, Score = 923, P = 1.1e-92

TREMBL:CEC38D4_5 gene: "C38D4.8"; Caenorhabditis elegans cosmid C38D4, N = 1, Score = 418, P = 3.6e-39

PIR:S66337 ADP-ribosylation factor 1 - Chlamydomonas reinhardtii, N = 1, Score = 373, P = 2.1e-34

SWISSPROT:ARF1_CHLRE ADP-RIBOSYLATION FACTOR 1., N = 1, Score = 372, P = 2.7e-34

>TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds.
Length = 186

HSPs:

Score = 923 (138.5 bits), Expect = 1.1e-92, P = 1.1e-92
Identities = 178/186 (95%), Positives = 184/186 (98%)

```
Query:      1 MGLLDRLSVLLGLKKKEVHVLCLGLDMSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS 60
            MGLLDRLS LLGLKKKEVHVLCLGLDMSGKTTIINKLKPSNAQSQ+I+PTIGFSIEKFKS
Sbjct:      1 MGLLDRLSGLLGLKKKEVHVLCLGLDMSGKTTIINKLKPSNAQSQDIVPTIGFSIEKFKS 60

Query:      61 SSSLFTVFDMMSGQGRYRNLEWHYYKEGQAIIFVIDSSDLRMVVAKEELDTLLNHPDIKH 120
            SSSLFTVFDMMSGQGRYRNLEWHYYK+GQAIIFVIDSSD+LRMVVAKEELDTLLNHPDIKH
Sbjct:      61 SSSLFTVFDMMSGQGRYRNLEWHYYKDGQAIIFVIDSSDKLRMVVAKEELDTLLNHPDIKH 120

Query:      121 RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGVLDWLQDQ 180
            RRIPILFFANKMDLRD+VTSVKVSQLLCLE+IKDKPWHICASDAIKGEGVLDWLQDQ
Sbjct:      121 RRIPILFFANKMDLRDSVTSVKVSQLLCLESIKDKPWHICASDAIKGEGVLDWLQDQ 180

Query:      181 IQTVKT 186
            IQ VKT
Sbjct:      181 IQAVKT 186
```

Pedant information for DKFZphtes3_23111, frame 3

Report for DKFZphtes3_23111.3

```
[LENGTH]      186
[MW]           21097.69
[pI]           8.72
[HOMOL]        TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog
ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds. 4e-94

[FUNCAT]       30.08 organization of golgi [S. cerevisiae, YDL192w] 1e-36
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 1e-36
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL192w]
1e-36
[FUNCAT]       30.09 organization of intracellular transport vesicles [S. cerevisiae,
YDL137w] 2e-36
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YBR164c] 2e-32
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 2e-32
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 4e-19
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 4e-19
[FUNCAT]       r general function prediction [M. jannaschii, MJ1339] 2e-05
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       10.05.07 g-proteins [S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YKR014c] 2e-04
[FUNCAT]       08.19 cellular import [S. cerevisiae, YKR014c] 2e-04
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YKR014c]
2e-04
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w]
4e-04
[BLOCKS]       BL01288C
[BLOCKS]       BL01020C SAR1 family proteins
[BLOCKS]       BL01019C ADP-ribosylation factors family proteins
```



```

SEQ      MGLDLRLSVLLGLKKKEVHVLCGLDNSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS
SEG      ..xxxxxxxxxxxxx.....
1hurA    .....CCCCEEEEETTTTCHHHHHHHHCCCCEEE--EETTEEEEEEE
SEQ      SLSFTVFDMSGQGRYRNLEWHYYKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH
SEG      .....
1hurA    TTEEEEEETTTTTTCCCHHHHHHCEEEEEEEETTTTTHHHHHHHHHHHHHHTTT--
SEQ      RRPIPIFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGVQEGVDWLQDQ
SEG      .....
1hurA    TTEEEEEETTTTTTCCCHHHHHHHCGGGTTTCEEEECBTTTTBTHHHHHHHHHHH
SEQ      IQTVKT
SEG      .....
1hurA    HHHHC.

```

Prosite for DKFZphtes3_23111.3

PS00017 24->32 ATP GTP A PDOC00017

Pfam for DKFZphtes3_23111.3

HMM_NAME	ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)		
HMM	*GMGwfsIfrkMWGLWNKEMRIMLGLDNAGKTTILYMLKlgE..IVTTI		
		MG++ ++ ++GL +KE+++L LGLDN+GKTTI+++LK+ ++	
Query	1	-MGLDLRLSVLLGLKKKEVHVLCGLDNSGKTTIINKLKPSNAQSQNIL	48
HMM	PTIGFNvetVeYkNIkFNvWDVGQdsIRPYWRHYYPNTDGIIWVVDsAd		
		PTIGF +E+ + ++F+V+D GQ + R +W HYY + ++II+V+DS+D	
Query	49	PTIGFSIEKfKSSLSFTVFDMSGQGRYRNlWEHYKQEGAIIFVVIDSSD	98
HMM	RDRMeEaKqELHaMLNEEEL..rDAPILiFANKQDLpGAMsEsEIREaLG		
		R RM AK+EL+ +LN+ ++ R+ P+L FANK DL++A+++ +++ +L	
Query	99	RLRMVVAKEELDTLLNHDPDKHRRIPILFFANKMDLRDAVTSVKVSQLLC	148
HMM	LHeIRcNrPwYIqMcCAvtGEGlyEGMDwLSnyInkrKk*		
		L++I+ + PWiI+ +++A++GEGl+EG DWL ++i+ K	
Query	149	LENiK-DKpWHICASDAIKGEGlQEGDwLDQIQTVKT	186

DKFZphtes3_23n19

group: testes derived

DKFZphtes3_23n19 encodes a novel 387 amino acid protein with similarity to rat protein kinase C-interacting RBCC protein 1.

The novel protein contains not the RING-B box-coiled coil (RBCC) motif of RBCC protein 1, and thus is not a member of this subgroup of RING finger proteins.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp

Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
1 CGGAGACCTT CGGGCCGTGT CCATTGTGG GCAAAGCCAG CGGGGCAGGC
51 TTGGCCAGAG TGCACCACTC GGGCCCGTCC CAGGCCCGAC GCTCTGGGCG
101 CGCCCGGAAC CCCAGGTTCC CGGCCCGTGT TTCCGACCCG CGGAGGGGGC
151 TCAGCGGGCCC GATCCACCGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
201 GACCGGAGAT GGGCCCGCCA GCGGGCGGGG CGGCGCGCGC GGCCTCGGAC
251 TTGGGCTCCG CCGCAGTGCT CTTGGCTGTG CACGCCCGCG TGAGGCCGCT
301 GGGCGCCGGG CCAGACGCGG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
351 CGGACCTTGA GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
401 CCTGGGGCGG TTAATTTGGA GTGGCCCTCG GAGTCAGTTT CCTACACCAT
451 CCGAGGGCCC ACCCAGCACG AGCTACAGCC TCCACCAGGA GGGCCTGGAA
501 CCCTCAGCCT GCACTTCCTC AACCTCAGG AAGCTCAGCG GTGGGCAGTC
551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
601 ACCACCAGCC TTGGGCCCAG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
651 CGGAAGCCTC CACACTCAAG GGCCCTCCAC CTGAGGCAGA TCTTCCTAGG
701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
801 TGGCCAGACA TCGTGTGGCC CTGAGTGTTC AGCTTCAGGA GGCCTGCTTC
851 CCACCTGGCC CCATCAGGCT GCAGGTCACA CTTGAAGACG CTGCCTCTGC
901 CGCATCCGCC GCGTCCTCTG CACACGTTGC CCTGCAGGTC CACCCCACT
951 GCACTGTTGC AGCTCTCCAG GAGCAGGTGT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCGTGTGT TGCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCCGGCAGGA TGGGGACCCG GCTTTCTCT
1101 ACTTGCTGTC AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCCAGA AGATGGACGG GGAACCTGGA CGCTTGTTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCCAGCCTG CCCAGTCCAC
1251 TCCAGCCAG CTGGTCCTGT CTTCTCTGCA CTTTCATCAA TGCCCCAGAC
1301 CGCCCTGGCT GTGAGATGTG TAGCACCAG AGGCCCTGCA CTTGGGACCC
1351 CCTTGCTGCA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCCTTCCC TCACAAGTCC GACATCTCCA GGCCCCCACT GAACTCCGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGGTTG GGGGGAAGGG
1501 CCACAAAATG AAACCATTA AGACCTTAA GAGCCAAAAA AAAAAAAAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387

Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAAEQ LRRQLSADP
51 ERPGFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTLS
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLLSPPEA
151 STLKGPPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEK AAQVAAVLAQ
201 HRVALSVQLQ EACFPPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRO DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLFPFS LGLPPGPQPA ASSLPSPLOP
351 SWSCPSCTFI NAPDRPGCEM CSTQRPCPWD PLAAAST

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1,
 Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
 complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
 HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
 = 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
 musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
 = 367, P = 9.3e-34

>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
 UbcM4 interacting protein 28 mRNA, complete cds.
 Length = 498

HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

```

Query: 175 LAGSLARAIAGGDEKGAQVAQVAHLAQHRVALSVQLQEQEACFPPIRLQVTLEDAASAASA 234
      +A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA
Sbjct: 1 MALSLARAVAGGDEQAIAKYATWLAERQVPLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQVFSELGFPFPAVQRWVIGRCLCVPERSLASYGVRODGD 294
      + + L V P TVA+L++ VF + GPPP++QWV+G+ L + +L S+G+R++GD
Sbjct: 57 -HTVTIWLTVRPDMTVASLKMVFLDYGFPPSLQWVVGQRLARDQETLHSHGIRNRDGD 115

Query: 295 AFLYLLSAPREAPATGSPQHPQK-----MDGELG--RLFPFSLG-LPPG-PQPAASSLP 345
      A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P
Sbjct: 116 AYLYLLSARN----TSLNPQELQQRQLRMLLEDLGFKDLTLQSRGPLEPVLKPRTNQEP 171

Query: 346 -----SPLQP--SWSCPSCTFINAPDRPGCEMCSTQRPCWT 379
      +P P W CP CTFIN P RPGCEMC RP T+
Sbjct: 172 GQPDAAPEPPVGWQCPGCTFINKPTRPGCEMCRRARPETY 212

```

Pedant information for DKFZphtes3_23n19, frame 2

Report for DKFZphtes3_23n19.2

```

[LENGTH]      387
[MW]           39949.29
[pI]           5.53
[HOMOL]        TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]       BL005788
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      17.57 %

```

```

SEQ  MAPPAGGAAAASDLGSAAVLLAVHAAVRPLGAGPDAAEQ LRRQLSADPERPGFRLEL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```


738

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387
 Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAAEQ LRLQLSADP
51 ERPGFRFREL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTSL
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPEEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLAQ
201 HRVALSVQLQ EACFPFPIR LQVTLDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLFPFS LGLPPGPQPA ASSLPSPLOP
351 SWSCPCTFI NAPDRPGCEM CSTQRPCTWD PLAAAST

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1,
 Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
 complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
 HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
 = 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
 musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
 = 367, P = 9.3e-34

>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
 UbcM4 interacting protein 28 mRNA, complete cds.
 Length = 498

HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

```

Query: 175 LAGSLARAIAGGDEKGAQAQVAQVLAQHRVALSVQLQEACFPFPIRLQVTLDAASAASA 234
      +A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA
Sbjct: 1 MALSLARAVAGGDEQAQAIKYATWLAEQVRVLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGD 294
      + + L V P TVA+L++ VF + GFPP++Q+WV+G+ L + +L S+G+R++GD
Sbjct: 57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQVWVGQRLARDQETLHSHGIRRNQDG 115

Query: 295 AFLYLLSAPREAPATGSPQHPQK-----MDGELG--RLFPPSLG-LPPG-PQPAASSLP 345
      A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P
Sbjct: 116 AYLYLLSARN----TSLNPQELQRQRLRMLDLGFKDLTQSRGPLEPVLPKPRTNQEP 171

Query: 346 -----SPLQP--SWSCPCTFINAPDRPGCEMCSTQRPCTW 379
      +P P W CP CTFIN P RPGCEMC RP T+
Sbjct: 172 GQPDAAPEPPVVGWQCPGCTFINKPTRPGCEMCRRARPETY 212

```

Pedant information for DKFZphtes3_23n19, frame 2

.....

```
[LENGTH]      387
[MW]          39949.29
[pI]          5.53
[HOMOL]       TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]      BL00578B
[KW]          Alpha_Beta
[KW]          LOW COMPLEXITY      17.57 %
```

```
SEQ MAPPAGGAAAAASDLGSAVLLAVHAARVLPGAGPDAAQLRRLQLSADPERPGRFLRL
SEG .XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD cccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhcccccceeee
```

```

SEQ      LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPGGPGTLSLHFLNPQEAQRWAVLVRGAT
SEG      .....
PRD      cccccccceccccceeeeeccccccccccccccccccccceeeeeccccchhhhhheeeccce

```

```
SEQ  VEQQNGSKSNSPPALGPEACPVSLSPPEASTLKGPPEADLPRSPGNLTEREELAGSLA
SEG  .....
PRD  eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeehhhhhhhhhhhhhhhhh
```

```
SEQ      RAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPFGPIRLQVTLEDAASAASAASSAHV
SEG      .....xxxxxxxxxxxxx..
PRD      hhhhhccccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccceecchhhhhhhhhhhhee
```

```

SEQ      ALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDPAFLYLL
SEG      .....
PRD      eeeeeeechhhhhhhhhhhhhccccccchhhhhhhhhhhhhccccccccccccccccceeeec

```

```

SEQ      SAPREAPATGSPQHFPQKMDGELGRLFFPSLGLPFGQPAASSLPSLQPSWSQPSCTFI
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...
PRD      cccccccccchhhhhhhhhhhccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      NAPDRPGCEMCSTQRPCTWDPLAAAST
SEG      .....
PRD      cccccccccccccccccccccceeeccc

```

(No Prosite data available for DKFZphtes3_23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

DKFZphtes3_26g22

group: intracellular transport/trafficking

DKFZphtes3_26g22 encodes a novel 898 amino acid protein with similarity to kinesins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport. It is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain contains a large globular N-terminal domain which is responsible for the motor activity of kinesin, which is known to hydrolyze ATP and to bind and move on microtubules. Several proteins involved in chromosome segregation and cell division contain this motor domain, such as drosophila claret segregational protein (ncd), Drosophila kinesin-like protein (nod), human CENP-E and human mitotic kinesin-like protein-1 (MKLP-1). The novel protein is a new kinesin like prottein.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

strong similarity to kinesins

Sequenced by EMBL

Locus: unknown

Insert length: 3032 bp

No poly A stretch found, no polyadenylation signal found

```
1 CTGAAGCGCT GGGAGGCCGA CATTAAAGTG AAGTGGTTGC GGTAACCTGG
51 CCTGGGCGCTG AAGTGAGTGA GAGGCACATG AAGAGAAGTA TTCAAGTATT
101 TATACAGATA GGAATCAAGA TAATCAACAA TGCTGTGCAC TGAGGAAGAC
151 CTGTGGCCACC ATATGAAAGT AGTAGTTCGT GTACGTCCGG AAAACACTAA
201 AGAAAAAGCA GCTGGATTTC ATAAAGTGGT TCATGTTGTG GATAAACATA
251 TCCTAGTTTT TGATCCCAAA CAAGAAGAAG TCAGTTTTTT CCATGGAAAG
301 AAAACTACAA ATCAAAATGT TATAAAGAAA CAAAATAAGG ATCTTAAATT
351 TGTATTTGAT GCTGTTTTTG ATGAAACGTC AACTCAGTCA GAAGTTTTTG
401 AACACACTAC TAAGCCAATT CTTCGTAGTT TTTTGAATGG ATATAATTGC
451 ACAGTACTTG CCTATGGTGC CACTGGTGCT GGGGAAGACCC ACACTATGCT
501 AGGATCAGCT GATGAACCTG GAGTGATGTA TCTAACAATG TTACACCTTT
551 ACAAATGCAT GGATGAGATT AAAGAAGAGA AAATATGTAG TACTGCAGTT
601 TCATATCTGG AGGTATATAA TGAACAGATT CGTGATCTCT TAGTAAATTC
651 AGGGCCACTT GCTGTCCGGG AAGATACCCA AAAAGGGGTG GTCGTTTCATG
701 GACTTACTTT ACACCAGCCC AAATCCTCAG AAGAAATTTT ACATTTATTG
751 GATAATGGAA ACAAACACAG GACACAACAT CCCACTGATA TGAATGCCAC
801 ATCTTCTCGT TCTCATGCTG TTTTCCAAAT TTACTTGCGA CAACAAGACA
851 AAACAGCAGG TATCAATCAA AATGTCCGTA TTGCCAAGAT GTCACCTATT
901 GACCTGGCAG GATCTGAGCG AGCAAGTACT TCCGGTGCTA AGGGGACCCG
951 ATTTGTAGAA GGCACAAATA TTAATAGATC ACTTTAGTCT CTTGGGAATG
1001 TCATCAATGC CTTAGCAGAT TCAAAGAGAA AGAATCAGCA TATCCCTTAC
1051 AGAAATAGTA AGCTTACTCG CTGTGTTAAAG GATTCTCTTG GAGGAAACTG
1101 TCAAACATAA ATGATAGCTG CTGTTAGTCC TTCCTCTGTA TTCTACGATG
1151 ACACATATAA CACTCTTAAG TATGCTAACC GGGCAAGGGA CATTAAATCT
1201 TCTTTGAAGA GCAATGTTCT TAATGTCAAT AATCATATAA CTCAATATGT
1251 AAAGATCTGT AATGAGCAGA AGGCAGAGAT TTTATTGTTA AAAGAAAAAC
1301 TAAAAGCCTA TGAAGAACAG AAAGCCTTCA CTAATGAAAA TGACCAAGCA
1351 AAGTTAATGA TTTCAAACCC TCAGGAAAAA GAAATCGAAA GGTTCACAGA
1401 AATCCTGAAC TGCTTGTTCC AGAATCGAGA AGAAATTAGA CAAGAATATC
1451 TGAAGTTGGA AATGTTACTT AAAGAAAATG AACTTAAATC ATTCTACCAA
1501 CAACAGTGCC ATAAACAAAT AGAAATGATG TGTTCTGAAG ACAAGTAGA
1551 AAAGGCCACT GGAAAACGAG ATCATAGACT TGCAATGTTG AAAACTCGTC
1601 GCTCCTACCT GGAGAAAAGG AGGGAGGAGG AATTGAAGCA ATTTGATGAG
1651 AATACTAATT GGCTCCATCG TGTCGAAAAA GAAATGGGAC TCTTAAGTCA
1701 AAACGGTCAT ATTCCAAAGG AACTCAAGAA AGATCTTCAT TGTCACCATT
1751 TGCACCTCCA GAACAAAGAT TTGAAAGCAC AAATTAGACA TATGATGGAT
1801 CTAGCTTGTC TTCAGGAACA GCAACACAGG CAGACTGAAG CAGTATTGAA
1851 TGCTTTACTT CCAACCTTAA GAAAAAATAA TTGCACATTA AAAGAAGCCG
1901 GCCTGTCAAA TGCTGCTTTT GAATCTGACT TCAAAGAGAT CGAACATTTG
1951 GTAGAGAGGA AAAAAGTGGT AGTTTGGGCT GACCAAACTG CCGAACAACC
2001 AAAGCAAAAC GATCTACCAG GGATTTCTGT TCTTATGACC TTTCCACAAC
2051 TTGGACCAAG TCAGCCTATT CCTTGTGCT CATCTTCAGG TGGAACTAAT
2101 CTGGTTAAGA TTCCTACAGA AAAAAGAACT CGGAGAAAAA TAATGCCATC
2151 TCCCTTGAAA GGACAGCATA CTCTAAAGTC TCCACCATCT CAAAGTGTGC
2201 AGCTCAATGA TTCTCTTAGC AAAGAAGCTT AGCCTATTGT ATATACACCA
2251 GAAGACTGTA GAAAAGCTTT TCAAAATCCG TCTACAGTAA CCTTAATGAA
2301 ACCATCATCA TTTACTACAA GTTTTCAGGC TATCAGTCA AACATAAACA
2351 GTGATAATTG TCTGAAAATG TTGTGTGAAG TAGCTATCCC TCATAATAGA
```



```

2401 AGAAAAGAAT GTGGACAGGA GGAAGTGGAC TCTACATTTA CTATATGTGA
2451 AGACATCAAG AGCTCGAAGT GTAAATTACC CGAACAAGAA TCACTACCAA
2501 ATGATAACAA AGACATTTTA CAACGGCTTG ATCCTTCTTC ATTCTCAACT
2551 AAGCATTCTA TGCCTGTACC AAGCATGGTG CCATCCTACA TGGCAATGAC
2601 TACTGTCTGC AAAAGGAAAC GGAAATTAAC AAGTTCTACA TCAAACAGTT
2651 CGTTAACTGC AGACGTAAAT TCTGGATTG CCAAACGTGT TCGACAAGAT
2701 AATTCAAGTG AGAAGCACTT ACAAGAAAAC AAACCAACAA TGGAACATAA
2751 AAGAAACATC TGTAAATAA ATCCAAGCAT GGTTAGAAAA TTTGGAAGAA
2801 ATATTTCAAA AGGAAATCTA AGATAAATCA CTTCAAAACC AAGCAAAATG
2851 AAGTTGATCA AATCTGCTTT TCAAAGTTTA TCAATACCCT TTCAAAAATA
2901 TATTTAAAAA CTTTGAAAGA AGACCCATCT TAAAGCTAAG TTTACCCAAG
2951 TACTTTGAGC AAGCAGAAAA ATGAAACTCT TTGTTTCTT CTTTGTGTGT
3001 CTAATAAAAT AAAATTTCAA AAGAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 130 bp to 2823 bp; peptide length: 898
 Category: strong similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (113-121)
 KINESIN_MOTOR_DOMAIN1 (252-264)

```

1 MSVTEEDLCH HMKVVVRVRP ENTKEKAAGF HKVVHVVDKH ILVFDPKQEE
51 VSFFHGKRTT NQNVIKKQNK DLKFVFDVAV DETSTQSEVF EHTTKPILRS
101 FLNGYNCVTL AYGATGAGKT HTMLGSADEP GVMYLTMLHL YKCMDEIKKEE
151 KICSTAVSYL EYVNEQIRDL LVNSGFLAVR EDTQKGVVVH GLTLHQPKSS
201 EEILHLLDNG NKNRTQHPTD MNATSSRSHA VFQIYLRQOD KTASINQNVN
251 IAKMSLIDLA GSERASTSGA KGTRFVEGTN INRSLLALGN VINALADSKR
301 KNQHIPYRNS KLTRLKDSL GGNQCOTIMIA AVSPSSVFYD DTYNLTLYAN
351 RAKDIKSSLK SNVLNVNNHI TQYVKICNEQ KAEILLKKEK LKAYEEQKAF
401 TNENDQAKLM ISNPQEKEIE RFQEILNCLF QNREEIRQEY LKLEMLLKEN
451 ELKSFYQQQC HKQIEMMCSE DKVEKATGKR DHRLAMLKTR RSYLEKRREE
501 ELKQFDENTN WLHRVEKEMG LLSQNGHIPK ELKKDLHCHH LHLQNKDLKA
551 QIRHMDLAC LQEQQHROTE AVLNALLPTL RKQYCTLKEA GLSNAAFESD
601 FKEIEHLVER KVVVVWADOT AEQPKQNDLP GISVLMTFPQ LGPVQPIPC
651 SSSGGTNLVK IPTEKRTRRK LMPSPKGGQH TLKSPPSQSV QLNDLSLSEL
701 QPIVYTPEDC RKAQNPSTV TLMKPSSFTT SFQAISNIN SDNCLKMLCE
751 VAIPHNRREK CGQEDLDSTF TICEDIKSSK CKLPEQESLP NDNKDILQRL
801 DPSSFSTKHS MPVPSMVPST MAMTTAAKRR RKLTSSTSNS SLTADVNSGF
851 AKRVRQDNSS EKHLENKPT MEHKNRICKI NPSMVRKFRG NISKGNLR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_26g22, frame 1

SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13., N = 3,
 Score = 874, P = 9e-93

TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila
 melanogaster kinesin like protein 67a mRNA, complete cds., N = 1, Score
 = 880, P = 4.2e-88

TREMBL:SPBC649_1 gene: "SPBC649.01c"; product: "putative kinesin-like
 protein"; S.pombe chromosome II cosmid c649., N = 3, Score = 814, P =
 9.8e-86

PIR:S64238 kinesin-related protein KIP3 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 802, P = 2.5e-83

>TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila

melanogaster kinesin like protein 67a mRNA, complete cds.
Length = 814

HSPs:

Score = 880 (132.0 bits), Expect = 4.2e-88, P = 4.2e-88
Identities = 181/345 (52%), Positives = 238/345 (68%)

```

Query:   11  HMKVVVRVRPENTKEKAAGFHKVVHVVDKHLVFDPKQEEVSFF-HGKKTNTQNVIKKQN 69
          ++KV VRVRP N +E          ++ V+D+ L+FDP +E+ FF G K   +++ K+ N
Sbjct:   8  NIKVAVRVRPYNVRELEQKQRSIIKVMDSALLFDPDEEDDEFFQGAQPYRDITKRMN 67

Query:   70  KDLKFVFDVAFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGATGAGKTHTMLGSADE 129
          K L  FD VFD ++ ++FE T P++ + LNGYNC+V YGATGAGKT TMLGS
Sbjct:   68  KKLTFEFDVFDIDNSNQLFEECTAPLVDVAVLNGYNCVSVFYGATGAGKTFTMLGSEAH 127

Query:   130  PGVMYLTMLHLKYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGFLAVREDTQKGVVV 189
          PG+ YLTM L+ + + + + VSYLEVYNE + +LL SGPL +RED GVVV
Sbjct:   128  PGLTYLTMQDLFDKIQAQSDVRKFDVGVSYLEVYNEHVMMNLLTKSGPLKLRDNN-GVVV 186

Query:   190  HGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSRAVFIYLRQQDKTASINQNV 249
          GL L  S+EE+L +L GN +RTQHPTD NA SRSRA+VQ+++R ++ + V
Sbjct:   187  SGLCLTPIYSAEELLRMLMLGNSHRTQHPTDANAESSRSRAIFQVHIRITERKTDTKRTV 246

Query:   250  RIAKMSLIDLAGSERASTSGAGKTRFVEGTNINRSLALGNVINALADSKRKNQHPIRYN 309
          K+S+IDLAGSERA+++ G RF EG +IN+SLALGN IN LAD + HIPYR+
Sbjct:   247  ---KLSMIDLAGSERAASTKGIGVRFKEGASINKSLLALGNCINKLADGLK---HIPYRD 300

Query:   310  SKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDI 355
          S LTR+LKDSLGGNC+T+M+A VS SS+ Y+DTYNTLKYA+RAK I
Sbjct:   301  SNLTRILKDSLGGNCRTLMVANVSMSSLTIEDTYNTLKYASRAKKI 346

```

Pedant information for DKFZphtes3_26g22, frame 1

Report for DKFZphtes3_26g22.1

```

[LENGTH]      898
[MW]           102281.63
[pI]           9.09
[HOMOL]        SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13. 3e-97
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]       08.22 cytoskeleton-dependent transport [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]       09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       03.13 meiosis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       11.01 stress response [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
                [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       30.05 organization of centrosome [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YKL079w]
4e-28
[BLOCKS]       BL00411H
[BLOCKS]       BL00411G
[BLOCKS]       BL00411F
[BLOCKS]       BL00411E Kinesin motor domain proteins
[BLOCKS]       BL00411C Kinesin motor domain proteins
[BLOCKS]       BL00411B Kinesin motor domain proteins
[BLOCKS]       BL00411A Kinesin motor domain proteins
[SCOP]         d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 1e-117
[SCOP]         d3kar_ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 1e-112
[PIRKW]        nucleus 6e-87
[PIRKW]        heterodimer 4e-68
[PIRKW]        DNA binding 9e-60
[PIRKW]        heterotetramer 2e-54
[PIRKW]        mitosis 9e-60
[PIRKW]        microtubule binding 4e-68
[PIRKW]        ATP 6e-87
[PIRKW]        phosphoprotein 5e-59
[PIRKW]        heterotrimer 4e-68
[PIRKW]        purine nucleotide binding 1e-26
[PIRKW]        P-loop 6e-87
[PIRKW]        coiled coil 4e-68
[PIRKW]        heptad repeat 3e-62
[PIRKW]        methylated amino acid 2e-54
[PIRKW]        hydrolase 2e-54
[PIRKW]        GTP binding 1e-60

```


[PIRKW] cell division 5e-57
 [SUPFAM] kinesin-related protein KIP1 3e-50
 [SUPFAM] kinesin-related protein CIN8 7e-33
 [SUPFAM] kinesin heavy chain 2e-54
 [SUPFAM] suppressor protein SMY1 1e-26
 [SUPFAM] kinesin-related protein KIF3 4e-68
 [SUPFAM] kinesin-related protein KIF2 1e-46
 [SUPFAM] kinesin-related protein unc-104 7e-60
 [SUPFAM] unassigned kinesin-related proteins 6e-87
 [SUPFAM] centromere protein E 3e-54
 [SUPFAM] kinesin-related protein KLP61F 5e-57
 [SUPFAM] kinesin-related protein MKLP-1 2e-28
 [SUPFAM] pleckstrin repeat homology 7e-60
 [SUPFAM] kinesin-related protein KIF1B 4e-61
 [SUPFAM] kinesin motor domain homology 6e-87
 [SUPFAM] kinesin-related protein KLP4 1e-43
 [SUPFAM] kinesin-related protein nodA 1e-30
 [SUPFAM] kinesin-related protein Eg5 5e-59
 [PROSITE] ATP_GTP_A 1
 [PROSITE] KINESIN_MOTOR_DOMAIN1 1
 [PFAM] Kinesin_motor_domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 8.57

SEQ MSVTEEDLCHHMKVVVRVPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKT
 SEG
 3kar-TBEEE

 SEQ NQNVIKKQNKDLKFVDAVDETSTQSEVFEHTTKPILRSFLNGYNCTVLYAGATGAGKT
 SEG
 3kar- EEEETTTTTEEEEEETEEETTTTCHHHHHHHHHH-HHHGGGGCCCCEEEEEECTTTTCHH

 SEQ HTMLGSADEPGVMYLTMLHLKYCMDEIKEEKICSTAVSYLEVNEQIRDLLVNSGFLAVR
 SEG
 3kar- HHHHTTTT--THHHHHHHHHHHHHHHGGGCEEEEEEEEEETEEET--TCCCCEE

 SEQ EDTQKGVVHGLTLHQPKSSEELHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQD
 SEG
 3kar- EETTTTEEEETTCCEECCGGHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEEEE

 SEQ KTASINQNVRIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLLAGNVINALADSKR
 SEG
 3kar- TTTTCEE---EEEEEEECCECCCCCCC---HHHHHHHHHHHHHHHHHHHHHHHTTTT

 SEQ KNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKSSLK
 SEG
 3kar- TTTCTTTTHHHHHHHGGGCTTTTEEEEEEECCCGGHHHHHHHHHHHH.....

 SEQ SNVLNVNHHITQYVKICNEQKAEILLKEKLKAYEEQKFTNENDQAKLMISNPQEKEIE
 SEG
 3kar-

 SEQ RFQEILNCLFQNREEIRQEYLLKLEMLLKENELKSFYQQQCHKQIEMMCSEDKVEKATGKR
 SEG
 3kar-

 SEQ DHRLAMLKTRRSYLEKRREEELKQFDENTNWLHRVEKEMGLLSQNGHIPKELKKDLHCHH
 SEG
 3kar-

 SEQ LHLQNKDLKAQIRHMDLACLQEQHRQTEAVLNALLPTLRKQYCTLKEAGLSNAAFESD
 SEG xxx.....
 3kar-

 SEQ FKEIEHLVERKKVVWADQTAEQPKQNDLPGISVLMTFPQLGPVQPIPCSSSSGGTNLVK
 SEG
 3kar-

 SEQ IPTEKRTRRLMPSPKLGQHTLKSPPSQSVQLNDSLSEKELQPIVYTPEDCRKAFQNPSTV
 SEG
 3kar-

 SEQ TLMKPSSFTTSFQAISSNINSNCLKMLCEVAIPHNRREKCGQEDLDSTFTICEDIKSSK
 SEG
 3kar-

 SEQ CKLPEQESLPNDNKDILQRLDPSSFSTKHSMPVPSMVPSPYAMTTAAKRKRKLTSSSNS
 SEG
 3kar-

SEQ SLTADVNSGFAKRVQRDQNSSEKHLQENKPTMEHKRNICKINPSMVRKFGRNISKGNLR
 SEG xxx.....
 3kar-

Prosites for DKFZphtes3_26g22.1

PS00017 113->121 ATP_GTP_A PDOC00017
 PS00411 252->264 KINESIN_MOTOR_DOMAIN1 PDOC00343

Pfam for DKFZphtes3_26g22.1

HMM_NAME Kinesin motor domain

HMM *RCRPLNeREindgcscvVQWPpWtGyktvhngheds.....
 R+RP N +E+++G +VV + + + + +++E S

Query 17 RVRPENTKEKAAGFHKVVHVVD-KHILVFDPKQEEVSFFHGKKTNNQNV 64

HMMphksFtFDHVFwWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQ
 + F+ED VF+ ++TQ +V++ + PI+ ++++GYNCT++AYG

Query 65 IKKQNKDLKFVDAVFDETSTQSEVFHTTKPILRSFLNGYNCTVLAYGA 114

HMM TGSGKTYTMMGpggehPDHmGIIPRcCHDIFdrIdkfgekDhdFWHvKCS
 TG+GKT+TM G + D+ G+ + +++++ D + + + +S

Query 115 TGAGKTHTMLG----SADEPGVMYLMHLHYKCMDEIK-EEKIC-STAVS 158

HMM YMEIYNEeIYDLLCPnPqhMkpLnIHEHPNMGPYVqGCTEfHVCsYeDac
 Y+E+YNE+I+DLL+ N ++PL+++E+ G+ V G+T+ +S E+++

Query 159 YLEVYNEQIRDLLV-N---SGPLAVREDTQKGVVHGLTLHQPKSSEEIL 204

HMM hWIWqGnknRHVAaTnMndhSSRShtIFTIHVeQrHk..qcdehvcHskM
 H+++ GNKNR+ +T MN++SSRSH++F+I ++Q K + V++ KM

Query 205 HLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASINQNVRIAKM 254

HMM NLVDLAGSERVnrTGAEGQRlKEGcNINqSLttLGnVinaLaDgqTKYmY
 +L+DLAAGER++ +GA G+R+ EG+NIN+SL++LGNVINALAD +

Query 255 SLIDLASERASTSGAKGTRFVEGTNINRSLALGNVINALADSK----- 299

HMM gghgHIPYRDSKLTWILQDSLGGNCkTcMIACIWPadWNYEETLSTLRYA
 +++HIPYR SKLT+LL+DSLGGNC T MIA+++P+ + Y++T +TL+YA

Query 300 RKNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYA 349

HMM dRAKnIkNkPQINEDPcamalWRrYheQIqdMKhql*
 +RAK+IK + N + + + +Y + + K++

Query 350 NRAKDIKSSLKSNVLNVN-NHITQYVKICNEQKAEI 384

DKFZphtes3_27d1

group: metabolism

DKFZphtes3_27d1 encodes a novel 712 amino acid protein similar to ubiquitin-specific proteases (EC 3.1.2.15).

The novel protein contains both, a ubiquitin carboxyl-terminal hydrolases family 2 signature 1 and signature 2. Pfam predicts a new member of the ubiquitin carboxyl-terminal hydrolases family 2. The ubiquitin system is responsible for the turn over of proteins. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The novel protein is a new member of the ubiquitin carboxyl-terminal hydrolases family 2, represented by proteins such as yeast UBP1-16, human tre-2, human isopeptidase T and others.

The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

similarity to ubiquitin-specific proteases

complete cDNA, complete cds, 4 EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2871 bp

Poly A stretch at pos. 2836, no polyadenylation signal found

```

1  CCAAACCTGA AAGAGGTTGA TTTGTAATGA TTTGCAGGGG GGCCTGGAG
51  GCAGCGGCCA GGACTTTTCA CTTAGGAGAT CAGCATTGCG CCTGATGGAA
101 ACTGGGCGAT CCTGCAGGGA CTGACCTCTG AGTTATCCAA AGGCCGACCT
151 GGGGAAAGAC TGATTTTGAG GTTTTAATAG TTTTCAGATG CTTCAAGTGT
201 TGTGAACAGA GACTTGTTTG GATTATGCAT TTCTCAGCTA GACTAAATAA
251 ATGCTAGCAA TGGATACGTG CAAACATGTT GGGCAGCTGC AGCTTGCTCA
301 AGACCATTCG AGCCTCAACC CTCAGAAATG GCACTGTGTG GACTGCAACA
351 CGACCGAGTC CATTGGGGCT TGCCTTAGCT GCTCCCATGT TGCCTGTGGA
401 AGATATATTG AAGAGCATGC ACTCAAGCAC TTTCAAGAAA GCAGTCATCC
451 TGTGTCATTG GAGGTGAATG AGATGTACGT TTTTGTGTAC CTTTGTGATG
501 ATTATGTTCT GAATGATAAC GCAACTGGAG ACCTGAAGTT ACTACGACGT
551 ACATTAAGTG CCATCAAAAG TCAAAATTAT CACTGCACAA CTCGTTAGTG
601 GAGGTTTSTA CGGTCCATGG GTACAGGTGA TGATTCTTAT TTCTTACATG
651 ACGGTGCCCA ATCTCTGCTT CAAAGTGAAG ATCAACTGTA TACTGCTCTT
701 TGGCAGAGGA GAAGGATACT AATGGGTAAG ATCTTTCGAA CATGGTTTGA
751 ACAATCACCC ATTGGAAGAA AAAAGCAAGA AGAACCATTG CAGGAGAAAA
801 TAGTAGTAAA AAGAGAAGTA AAGAAAAGAC GGCAGGAATT GGAGTATCAA
851 GTTAAAGCAG AATTGGAAG TATGCCTCCA AGAAAGAGTT TACGTTTACA
901 AGGGCTCGCT CAGTCGACCA TAATAGAAAT AGTTTCTGTT CAGGTGCCAG
951 CACAAACGCC AGCATCAGCA GCAAAAGATA AAGTACTCTC TACCTCAGAA
1001 AATGAAATAT CTCAAAAAGT CAGTGACTCC TCAGTTAAAC GAAGGCCAAT
1051 AGTAACCTCT GGTGTAACAG GATTGAGAAA TTTGGGAAAT ACTTGCTATA
1101 TGAATCTCTG TCTTCAGGTG TTGAGTCATT TACTTATTTT TCGACAATGT
1151 TTTTAAAGC TTGATCTGAA CCAATGGCTG GCTATGACTG CTAGCGAGAA
1201 GACAAGATCT TGTAAGCATC CACCAGTCAC AGATACAGTA GTATATCAAA
1251 TGAATGAATG TCAGGAAAAA GATACAGGTT TTGTTTGCTC CAGACAATCA
1301 AGTCTGTCTA CAGGACTAAG TGGTGGAGCA TCAAAAGGTA GAAAGATGGA
1351 ACTTATTCAG CCAAAGGAGC CAACTTCACA GTACATTTCT CTTTGTCTATG
1401 AATTGCATAC TTTGTTCCAA GTCATGTGGT CTGGAAAGTG GGCCTTGGTC
1451 TCACCATTTG CTATGCTACA CTCAGTGTGG AGACTCATTC CTGCCTTTCTG
1501 TGGTTACGCC CAACAAGACG CTCAGGAATT TCTTTGTGAA CTTTGTAGATA
1551 AAATACAACG TGAATTAGAG ACAACTGGTA CCAGTTTACC AGCTCTTATC
1601 CCCACTTCTC AAAGGAAACT CATCAACAAA GTTCTGAATG TTGTAATAAA
1651 CATTTTTCAT GGACAATTC TTAGTCAGGT TACATGTCTT GCATGTGACA
1701 ACAAATCAAA TACCATAGAA CCTTCTGGG ACTTGTCATT GGAGTTTCCA
1751 GAAAGGTATC AATGCAGTGG AAAAGATATT GCTTCCCAGC CATGTCTGGT
1801 TACTGAAATG TTGGCCAAAT TTACAGAAAC TGAAGCTTTA GAAGGAAAAA
1851 TCTACGTATG TGACCAAGTG AACTCAAAGC GTAGAAGGTT TTCCTCCAAA
1901 CCAGTTGTAC TCACAGAAGC CCAGAAACAA CTTATGATAT GCCACCTACC
1951 TCAGGTTCTC AGACTGCACC TCAACAGATT CAGGTGGTCA GGACGTAAATA
2001 ACCGAGAGAA GATTGGTGT CATGTTGGCT TTGAGGAAAT CTTAAACATG
2051 GAGCCCTATT GCTGCAGGGA GACCTGAAA TCCCTCAGAC CAGAATGCTT
2101 TATCTATGAC TTGTCGCGG TGTTGATGCA CCATGGGAAA GGATTTGGCT
2151 CAGGGCAGTA CACTGCCTAC TGCTATAATT CTGAAGGAGG GTTCTGGGTA
2201 CACTGCAATG ATTCCAACT AAGCATGTGC ACTATGGATG AAGTATGCAA
2251 GGTCAAGCT TATATCTTGT TTTATACCCA ACGAGTACT GAGAATGGAC

```



```

2301 ATTCTAACT TTTGCCTCCA GAGCTCCTGT TGGGGAGCCA ACATCCCAAT
2351 GAAGACGCTG ATACCTCGTC TAATGAAATC CTTAGCTGAT CCAAAGACAA
2401 TGGGGTTTTT TCCCTGTGAT TTATATATAT ACTTTTTTAA AGACTGATGT
2451 ACCATTTTAA ACTTCATTTT TTCTTGTGAA TCAGTGTATA CTACATTTAT
2501 ACATTTTATA TCTAACAAAT TTTTTTTTTT ACAAAGTATA AATGTATATA
2551 TCAACTGAAG GTAACACTT TTTTCATATT TGGAGTTTGA AACTTTTGGT
2601 GTTTACCTCA GACTGATGTT ACCTCTTTTA TATTTTATG TCTTAATTGG
2651 CTCGGATGAT GAACCTGTGC AATCTTCTAC CAACAAAGTT CAAGTGCCAT
2701 CATTTTATAT ACATGTATCT TTTTCAGGTA TTTTCTATAC AAATTCTTAA
2751 TAGATGGAAG ATTAGACTCT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AAAAAAAAAA AAAAAAAAAA AAGGGCGCGC CGCTCTAAAA AAAAAAAAAA
2851 AAAAAAAAAA AAAAAAAAAA G

```

BLAST Results

No BLAST result

Medline entries

98072201:
Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.

98431658:
The ubiquitin system.

Peptide information for frame 2

ORF from 251 bp to 2386 bp; peptide length: 712
Category: similarity to known protein
Prosite motifs: UCH_2_1 (274-290)
UCH_2_2 (619-638)
UCH_2_2 (619-638)

```

1 MLAMDTCKHV GQLQLAQDHS SLNPQKWHCV DCNTTESIWA CLSCSHVACG
51 RYIEEHALKH QESSHHPVAL EVNEMYVFCY LCDDYVLNDN ATGDLKLLRR
101 TLSAIKSONY HCTTRSGRFL RSMGTGDDSY FLHDGAQSLL QSEDQLYTAL
151 WHRRRILMGK IFRTWFEQSP IGRKKQEEFF QEKIVVKREV KRRRQELEVO
201 VKAELESMPK RKSRLRQGLA QSTIIIEIVS QVPAQTPASP AKDKVLSTSE
251 NEISQKVSDD SVKRRPIVTP GVTGLRLNGN TCYMNSVLQV LSHLLIFRQC
301 FLKLDLNLWL AMTASEKTRS CKHPPVTDTV VYQMNCEQEK DTGFVCSROS
351 SLSSGLSGGA SKGRKMELIQ PKEPTSQYIS LCHELHTLFQ VMWSGKWALV
401 SPFAMLSHVW RLIPAFRGYA QDDAQEFLCE LLDKIQRELE TTGTSPLALI
451 PTSQRKLIKQ VLNVVNNIFH GQLLSQVTCL ACDNKSNTIE PFWDLSLEFP
501 ERYQCSGKDI ASQPCLVTEM LAKFETEAL EGKIYVCDQC NSKRRRFSSK
551 PVVLTEAQKQ LMICHLPPQL RLHLKRFRRS GRNNREKIGV HVGFEIILNM
601 EPYCCRETLK SLRPECFIYD LSAVVMHGGK GFGSGHYTAY CYNSEGGFWV
651 HCNDSKLSMC TMDEVCKAQA YILFYTQRTV ENGHSKLLPP ELLLGSOHPN
701 EDADTSSNEI LS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27d1, frame 2

PIR:S57591 hypothetical protein YMR223w - yeast (*Saccharomyces cerevisiae*), N = 4, Score = 218, P = 8.4e-38

SWISSPROT:UBPB HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (KIAA0055), N = 2, Score = 300, P = 9.3e-31

TREMBL:AF079565_1 gene: "Ubp41"; product: "ubiquitin-specific protease UBP41"; *Mus musculus* ubiquitin-specific protease UBP41 (Ubp41) mRNA, complete cds., N = 3, Score = 187, P = 8.7e-30

PIR:I58376 hypothetical protein unip - mouse, N = 3, Score = 214, P = 1.2e-28

>SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13)
 (DEUBIQUITINATING ENZYME 11) (KIAA0055).
 Length = 1,118

HSPs:

Score = 300 (45.0 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 95/301 (31%), Positives = 149/301 (49%)

Query: 381 LCHELHTLFQVMWSGKVALVSPFAMLHSVWRLIPAFRGYAQQDAQEFLCELLDKIQREL- 439
 + E + + +W+G++ +SP ++ ++ F GY+QD+QE L L+D + +L
 Sbjct: 826 VAEFGIIMKALWTGQYRISPKDFKITIGKINDQFAGYSQQSQELLLFLMDGLHEDLN 885

Query: 440 -----ETTGTSLPALIPTSQRKLIKQVLN--VVNNIFHGQLLSQVTCACDNKSNT 488
 E L + LN ++ +F GQ S V CL C KS T
 Sbjct: 886 KADNRKRYKEENNDHLDFFKAAEHAWQKHKLNESIIVALFQQGFKSTVQCLTCHKKSRT 945

Query: 489 IEPFWDLSEFFERYQCSGKDIAQPCLVTEMLAKFTETEALGKIYVCDQCNSKRRRFS 548
 E F LSL +C+ +D CL + +K E + + + C C ++R
 Sbjct: 946 FEAFMYLSLPLASTSKCTLQD-----CL--RLFSK--EEKLTDNNRFYCSHCRARR---- 992

Query: 549 SKPVVLTEAQKQLMICHLPQVLRHLKRFWRSGRNNREKIGVHVGFEEILNMEPYCC-- 605
 ++ K++ I LP VL +HLKRF + GR ++K+ V F E L++ Y
 Sbjct: 993 -----DSLKKIEIWKLPVLLVHLKRFSYDGRW-KQKLQTSVDFPLENLDLSQYVIGP 1044

Query: 606 RETLKLRLPECFIYDLSAVVMHMGKFGSGHYTAYCYNSEGGFWVHCNDSKLSMCTMDEV 665
 + LK Y+L +V H+G G GHYTAYC N+ W +D ++S ++ V
 Sbjct: 1045 KNNLKK-----YNLFSVSNHYG-GLDGGHYTAYCKNAARQWFKFDDEHVS DISVSSV 1096

Query: 666 CKAQAYILFYTQ---RVTE 681
 + AYILFYT RVT+
 Sbjct: 1097 KSSAAYILFYTSLGPRVTD 1115

Score = 126 (18.9 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 41/116 (35%), Positives = 63/116 (54%)

Query: 200 QVKAELSMPPR--KSLRLQGLAQSTIIEIVSVQVPAQTASPAPKDKVLSTSENEISQKV 257
 Q+ AE + P + +S + Q+ I+ + P TP ++K + EIS ++
 Sbjct: 701 QIPAERDREPSKLKRSYSSPDITQA--IQEEKRRKPTVTPTVNRENKPTCYPKAEIS-RL 757

Query: 258 SDSSVKR-RPIVT---PGVTGLRNLGNTCYMNSVLQVLS---HLLIF--RQCFLKLDLNQ 308
 S S ++ P+ P +TGLRNLGNTCYMNS+LQ L HL + R C+ D+N+
 Sbjct: 758 SASQIRNLNPVFGSGPALTGLRNLGNTCYMNSILQCLCNAPHLADYFNRNCYQD-DINR 816

Score = 50 (7.5 bits), Expect = 8.3e-23, Sum P(2) = 8.3e-23
 Identities = 29/106 (27%), Positives = 51/106 (48%)

Query: 173 RKKQEFPFQEKIVVKREVKKRRQLEYQVKAELSMPPRKSRLRQGLAQSTIIEIVSVQV 232
 + KQE+ +E+ +++ K R++E E + K + E+ + Q A+ + + S Q
 Sbjct: 475 KKKQEKLRLRQEEQKEKLKKEEQKAKKKQEA-EENEITEKQKAKEEMKKESEQA 533

Query: 233 PAQ---TPASPAKD---KVLSTSENEIS--QKVSDDSVKRRPIVTPGV 272
 + T A K+ K S SE+E S +K + KR P TP +
 Sbjct: 534 KKEDKETSARKGKEITGVKRQSKSEHETSDAKKSVEDRGKRC--TPEI 580

Score = 42 (6.3 bits), Expect = 5.7e-22, Sum P(2) = 5.7e-22
 Identities = 13/58 (22%), Positives = 27/58 (46%)

Query: 167 EQSPIGRKKQEFPFQEKIVVKREVKKRRQLEY-QVKAELSMPPRKSRLRQGLAQST 223
 EQ +KKQE E +++ K+ ++ E Q K E + ++ + G+ + +
 Sbjct: 498 EQEQKAKKKQEAENEITEKQKAKEEMKKESEQAKKEDKETSARKGKEITGVKRQS 555

Pedant information for DKFzphes3_27d1, frame 2

Report for DKFzphes3_27d1.2

[LENGTH] 712
 [MW] 81155.71
 [pI] 8.21
 [HOMOL] SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING
 ENZYME 11) (KIAA0055). 4e-32
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YMR223w] 5e-33
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YMR223w] 5e-33


```

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YBL067c] 3e-19
[FUNCAT] 10.03.99 other osmosensing activities [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL186w] 4e-17
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YHL010c] 3e-12
[BLOCKS] BL00970A Nuclear transition protein 2 proteins
[BLOCKS] BL00972D
[BLOCKS] BL00972C
[BLOCKS] BL00972B
[BLOCKS] BL00972A
[EC] 3.1.2.15 Ubiquitin thiolesterase 5e-06
[PIRKW] alternative splicing 2e-11
[PIRKW] thiolester hydrolase 5e-06
[PIRKW] hydrolase 1e-14
[SUPFAM] RING finger homology 7e-11
[SUPFAM] deubiquinating enzyme SSV7 5e-16
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 10
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] UCH_2_1
[PROSITE] PKC_PHOSPHO_SITE 17
[PROSITE] ASN_GLYCOSYLATION 4
[PROSITE] UCH_2_1_1
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 4.92 %

```

```

SEQ      MLAMDTCKHVQQLAQDHSSLNPQKWHVCDCNTTESIWACLSCSHVACGRYIEEHALKH
SEG
PRD      .ccccccccchhhhhhhccccccccceeeccceeeeeccccccccchhhhhhhhhh

```

```

SEQ      FQESSHPVALEVNEMYFCYLCDDYVLNDNATGDLKLLRRTLSAIKSQNYHCTTRSGRFL
SEG      .....
PRD      hhhhccceeeccccceeeeeccccccccccccchhhhhhhhhhhhhcccccecccccc

```

```
SEQ      RSMGTGDDSYFLHDGAQSLQSEDQLYTALWHRRRILMGKIFRTWFEQSPIGRKKQEFPF  
SEG      .....  
PRD      cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
SEQ      QEKIVVKREVKKRRQLELEYQVKAELSMPPRKSLRLQGQAQSTIIIEIVSVQPAPQTASP  
SEG      xxxxxxxxxxxxxxxxxxxx.....  
PRD      hheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccceeeeccccccccccc
```

```
SEQ      AKDKVLSTSENEISQKVDSSVKRRPIVTPGVTGLRNLGNTCYMNSVLQVLSHLLIFRCQ
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhh
```

```

SEQ      FLKLDLNQWLAMTASEKTRSCKHPPVTDTVVYQMNECQEKDTGFVCSRQSSLSGSLGGA
SEG
PRD      hhhhhhhchhhhhhhhhhhhhhhhhccccccceehhhhhcccccccccccccccccccccc

```

```

SEQ      SKGRKMELIQKPTSQYISLCHELHTLFQVMWSGKWALVSPFAMLHSVWRLIPAFRGYA
SEG      xxxxxx.....
PRD      cccceeeccccccccchhhhhhhhhhhhhhhhhccceeecccchhhhhhhhhhhccccch

```

SEQ QQDAQEFLCELLDKIQRELETTGTSPLALIPTSQRKLIKQVLNVVNNIFHGQLLSQVTCL
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhccccchhhhhhhc

```
SEQ      ACDNKSNTIEPFDLSLEFPERYQCSGKDIAQPCLVTEMLAKFTETEALGKIYVCDQC  
SEG      .....  
PRD      cccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhcccceecccc
```

SEQ NSKRRRFSSKPVVLTEAQQLMICHLPQVLRLLHKKRFRWSGRNNREKIGVHVGFEEILNM
SEG
PD cccccccscchhhhhhhhhhhhhhhcchhhhhhhhhhhhhccccccccceeeccccccc

```
SEQ      EPYCCRETLKSLRPECFIYDL SAVVMHHGKGFGSGHYTAYCYNSEGGFWVHCNDSKL SMC
SEG      .....
PRD      CCCCCCCCCCCCCCCEEEEEEEEEEEC CCCCCCCCCEEE EC CCCCCCEE EEC CCCCC
```

```
SEQ      TMDEVCKAQAYILFYTQRVTENGHSKLLPPELLLGSQHPNEDADTSSNEILS
SEG
PP      cchhhhhhhhhhhhhhhheeecccccccccccccccccccccccccccccccccc
```


Prosites for DKFZphtes3_27d1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00001	90->94	ASN_GLYCOSYLATION	PDOC00001
PS00001	484->488	ASN_GLYCOSYLATION	PDOC00001
PS00001	653->657	ASN_GLYCOSYLATION	PDOC00001
PS00004	545->549	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	506->509	PKC_PHOSPHO_SITE	PDOC00005
PS00005	542->545	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	580->583	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	611->614	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	223->227	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	525->529	CK2_PHOSPHO_SITE	PDOC00006
PS00006	661->665	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	193->200	TYR_PHOSPHO_SITE	PDOC00007
PS00007	192->200	TYR_PHOSPHO_SITE	PDOC00007
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	359->365	MYRISTYL	PDOC00008
PS00008	471->477	MYRISTYL	PDOC00008
PS00008	589->595	MYRISTYL	PDOC00008
PS00009	171->175	AMIDATION	PDOC00009
PS00009	362->366	AMIDATION	PDOC00009
PS00972	274->290	UCH_2_1	PDOC00750
PS00973	619->638	UCH_2_2	PDOC00750

Pfam for DKFZphtes3_27d1.2

HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*GIqNlGNTCYMNSIIQCL*		
	G++NLGNTCYMNS++Q+L		
Query	274	GLRNLGNTCYMNSVLQVL	291
HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*YdLYgVICHYGntldyGHYWayVKNenHHRWkWYYFDDEtV*		
	YDL +V+ H+G + ++GHY+AY++N + ++W+ +D++		
Query	619	YDLSAVVMHKGKFGSGHYTAYCYNSE--GGFWVHCNDSKL	657

DKFZphtes3_27k4

group: transmembrane protein

Summary DKFZphtes3_27k4 encodes a novel 490 amino acid protein with similarity to two hypothetical C.elegans proteins.

The novel protein contains 10 transmembrane regions and a leucine zipper. It is a member of the new 10 trans-membrane domain containing protein family which is specific for multicellular eukariotes.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

strong similarity to C.elegans K07H8.2/ZK185.2
membrane regions: 10

complete cDNA, complete cds potential start at Bp 109, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1901 bp
Poly A stretch at pos. 1866, no polyadenylation signal found

```
1 GTGATTTACC AGAAAAACCA AGAAGACAGG CACAAAAAAG CAAACGGCAT
51 TTGGCAAGAT GGATTATCAA CTGCAGTACA GACTTTTAGT AATAGATCTG
101 AGCAACACAT GGAGTATCAC AGTTTCTCAG AGCAGTCTTT TCATGCCAAT
151 AATGGGCACG CATCATCAAG CTGCAGCCAA AAGTATGATG ACTATGCCAA
201 TTATAATTAC TGTGATGGAA GGGAGACTTC AGAAACCACT GCCATGTTAC
251 AAGATGAAGA TATATCTAGT GATGGTGATG AAGATGCTAT TGTAGAAGTG
301 ACCCCAAAAT TACCAAAGGA ATCCAGTGGC ATCATGGCAT TGCAAATACT
351 TGTGCCCTTT TTGCTAGCTG GTTTTGGAAC AGTTTCAGCT GGCATGGTAC
401 TGGATATAGT ACAGCACTGG GAGGTGTTCA GAAAAGTTAC AGAAGTTTTC
451 ATTTTAGTCC CTGCACCTCT TGGTCTCAAA GGGAACTTGG AAATGACATT
501 GGCATCCAGA TTATCCACTG CAGTAAATAT TGGGAAGATG GATTCACCCA
551 TTGAAAAGTG GAACCTAATA ATTGGCAACT TGGCTTTAAA GCAGGTTTCAG
601 GCAACAGTAG TGGGTTTCTC AGCAGCTGTG GCAGCAATTA TATTGGGCTG
651 GATTCCAGAA GGAAAATATT ACCTTGATCA TTCCATACTT CTGTGCTCTA
701 GCAGTGTGGC AACTGCCTTC ATTGCATCTC TTCTGCAGGG AATAAATATG
751 GTTGGGGTTA TCGTTGGTTC AAAGAAGACT GGTATAAATC CTGATAATGT
801 TGCTACACCC ATTGCTGCTA GTTTTGCGCA CCTTATAACT CTTGCCATAT
851 TGGCTTGGAT AAGTCAGGGC TTATACTCCT GTCTTGAGAC CTATTACTAC
901 ATTTCTCCAT TAGTTGGTGT ATTTTCTCTG GCTCTAACCC CTATTGGAT
951 TATAATAGCT GCCAAACATC CAGCCACAAG AACAGTTCTC CACTCAGGCT
1001 GGGAGCCTGT CATAACAGCT ATGGTTATAA GTAGCATTGG GGGCCTTATT
1051 CTGGACACAA CTGTATCAGA CCCAAACTTG GTTGGGATTG TTGTTTACAC
1101 GCCAGTTATT AATGGTATTG GTGGTAATTT GGTGGCCATT CAGGCTAGCA
1151 GGATTCTTAC CTACCTCCAT TTACATAGCA TTCCAGGAGA ATTGCCTGAT
1201 GAACCCAAAG GTTGTTACTA CCCATTAGA ACTTCTTTG GTCCAGGAGT
1251 AAATAATAAG TCTGCTCAAG TTCTACTGCT TTTAGTGATT CCTGGACATT
1301 TAATTTTCCT CTACACTATT CATTTGATGA AAAGTGGTCA TACTTCTTTA
1351 ACTATAATCT TCATAGTAGT GTATTTATTT GCGCGTGTGT TACAGGTATT
1401 TACCTTGCTG TGGATTGCTG ACTGGATGGT CCATCACTTC TGGAGGAAAG
1451 GAAAGGACCC GGATAGTTTC TCCATCCCTC ACCTAACAGC ATTGGGTGAT
1501 CTGCTCGGGA CAGCTCTGTT AGCCTTAAGT TTTCATTTTC TTTGGCTTAT
1551 TGGAGATCGA GATGGAGATG TTGGAGACTA ATAAATTCTA CAAACTGCTC
1601 TCAAGTTACC AAGGAAGAAA ATACACGACA ACCACTTATC GCTCTTTTTC
1651 AAAACTCTTA AATCAGTAGT TTGACTTTTG CCAGGGTAAT CTTCAGTTGG
1701 CCCTGATTCA ATTAATGGC CTTAATTTT TTTTAAGGAA TTTGTGTCAA
1751 AACCAGAATG AAGAGTATTC GTGCTGCTTT TCATAGAATA AATGATAATT
1801 TGACATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAGGGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGG
1901 G
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 109 bp to 1578 bp; peptide length: 490
Category: similarity to unknown protein

```
1 MEYHSFSEQS FHANNHGASS SCSQKYDDYA NNYNYCDGRET SETTAMLQDE
51 DISSDGDDEDA IVEVTPKLPK ESSGIMALQI LVPFLLAGFG TVSAGMVLDI
101 VQHWEVFRKV TEVFILVPAL LGLKGNLEMT LASRLSTAVN IGMDSPIEK
151 WNLIIGNLAL KQVQATVVGFLA AVAAIILG WIPEGKYLD HSILLCSSSV
201 ATAFIASLLQ GIIMVGIVG SKKTGINPDN VATPIAASFG DLITLAILAW
251 ISQGLYSCL EYIIISPLVG VFFLALTPIW IIIAAKHPAT RTVLHSGWEP
301 VITAMVISSI GGLILDTTVS DPNLVGIVVY TPVINGIGGN LVAIQASRIS
351 TYLHLHSIPG ELPDEPKGCY YPFRFFGPG VNNKSAQVLL LLVIPGHLIF
401 LYTIHLMKSG HTSLTIIFIV VYLFQAVLQV FTLLWIADWM VHHFWRKGD
451 PDSFSIPYLT ALGDLGLTAL LALSFFHFLWL IGDRLGQDVG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27k4, frame 1

TREMBL:AF036704.2 gene: "ZK185.2"; *Caenorhabditis elegans* cosmid
ZK185., N = 1, Score = 730, P = 3.1e-72

TREMBL:AF047659.9 gene: "K07H8.2"; *Caenorhabditis elegans* cosmid
K07H8., N = 1, Score = 940, P = 1.7e-94

>TREMBL:AF047659.9 gene: "K07H8.2"; *Caenorhabditis elegans* cosmid K07H8.
Length = 507

HSPs:

Score = 940 (141.0 bits), Expect = 1.7e-94, P = 1.7e-94
Identities = 204/412 (49%), Positives = 271/412 (65%)

```
Query: 68 LPKESSGIMALQILVFPFLLAGFGTVSAGMVLDIVQHWEVFRKVTEVFILVPALLGLKGNL 127
      +P ESS ++ Q+L PF +AG G V AG+VL IV W +F ++ E+ ILVPALLGLKGNL
Sbjct: 82 IPAESSYVLFQVLFPPFAVAGLGMVFAGLVLSIVVTWPLFEEIPEILILVPALLGLKGNL 141

Query: 128 EMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALQVQATVVGFLA AVAAIILGWIPEGKY 187
      EMTLASRLST N+G MDS ++ +++I NLAL QVQATVV FLA+ A L +IP G +
Sbjct: 142 EMTLASRLSTLANLGHMDSSQKQKDVVIANLALVQVQATVVAFASAFAAALAFIPSGDF 201

Query: 188 YLDHSILLCSSSVATAFIASLLQGIIMVGIVGSKKTGINPDNVATPIAASFGDLITLAI 247
      H L+C+SS+ATA ASL+ ++MV VIV S+K INPDNVATPIAAS GDL TL +
Sbjct: 202 DWAHGALMCASSLATACASLVL SLLMVVIVTSRKNINPDNVATPIAASLGDLTTLTV 261

Query: 248 LAWISQGLYSCL EYIIISPLVG VFFLALTPIWIIIAAKHPATRTVLHSGWEPVITAMVI 307
      LA+ T +++ +V V FL L P WI IA ++ T+ L++GW PVI +M+I
Sbjct: 262 LAFFGSVFLKAHNTESWLNIVIVLFLLLLPFWIKIANENEGTQETLYNGWTPVIMSMIL 321

Query: 308 SSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPGELPDEPK 367
      SS GG IL+T V + + Y PV+NG+GGNL A+QASR+STY H G LP+E
Sbjct: 322 SSAGGFIL EAVRRYH--SLSTYGPVLNGVGGNLAAVQASRLSTYFHKAGTVGVLPNEWT 379

Query: 368 GCYYPF--RTFFGPGVNNKSAQVLLLVIPGHLIFLYTIHLM---KSGHTSLTIIFIVV 421
      + R FF +++SA+VLLLV+PGH+ F + I L K+ T +F +
Sbjct: 380 VSRFTSVQRAFFSKEDSR SARVLLLVVPGHICFNELIQLFTLT SKNNVTPHGPLFTSL 439

Query: 422 YLFQAVLQVFTLLWIADWMVHHFWRKGD PDSFSIPYLTALGDLGLTALLALS F 475
      Y+ A++QV LL++ +V W+ DPD+ IPYLTALGDLGLT LL + F
Sbjct: 440 YMIAAIIQVVILLFVCQLLVALLWKWKIDPDNSVIPYLTALGDLGLTGLLFIV F 493
```

Pedant information for DKFZphtes3_27k4, frame 1

Report for DKFZphtes3_27k4.1

[LENGTH] 490
[MW] 53266.39


```

SEQ      IGDRDGDVGD
SEG      .....
PRD      ecccccccc
MEM      MM.....

```

PS000001	383->387	ASN_GLYCOSYLATION	PDOC000001
PS000004	108->112	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	65->68	PKC_PHOSPHO_SITE	PDOC000005
PS000005	221->224	PKC_PHOSPHO_SITE	PDOC000005
PS000006	5->9	CK2_PHOSPHO_SITE	PDOC000006
PS000006	54->58	CK2_PHOSPHO_SITE	PDOC000006
PS000006	146->150	CK2_PHOSPHO_SITE	PDOC000006
PS000006	238->242	CK2_PHOSPHO_SITE	PDOC000006
PS000006	257->261	CK2_PHOSPHO_SITE	PDOC000006
PS000006	296->300	CK2_PHOSPHO_SITE	PDOC000006
PS000006	318->322	CK2_PHOSPHO_SITE	PDOC000006
PS000007	25->33	TYR_PHOSPHO_SITE	PDOC000007
PS000008	90->96	MYRISTYL	PDOC000008
PS000008	122->128	MYRISTYL	PDOC000008
PS000008	216->222	MYRISTYL	PDOC000008
PS000008	220->226	MYRISTYL	PDOC000008

PS00008	254->260	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00013	12->23	PROKAR_LIPOPROTEIN	PDOC00013
PS00013	248->259	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	459->481	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_27k4.1)

DKFZphtes3_27o14

group: testes derived

DKFZphtes3_27o14 encodes a novel 358 amino acid protein with similarity to *C. elegans* cosmid C55A6.

The new protein contains a C3HC4 zinc finger (RING finger) signature. The ring finger structure binds two atoms of zinc, and is involved in mediating protein-protein interactions. No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *C.elegans* C55A6.1

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: /map="6"

Insert length: 2158 bp

Poly A stretch at pos. 2137, polyadenylation signal at pos. 2120

```
1 CCGAGGCCAG AGAGAAAAGA CTGCCAGGTG GCCGCAGCTG TGGCCGGAGA
51 GCACAAAGAA TGAACCAGCA GTGGAAGAGA AAATACTGTA AGCTGGCTGA
101 CTGCTGGTGA AGAAAATGCT TTATTTTGT GGCAGGCATC TGTGGGATCT
151 GTAATAGAAA TATATTGGAG TAATTCAAGA TTCTGTGGTT GGGCCCTTTG
201 ACTGCTCTCT CTACAGGTTT AATTGGGCA TTTACTCATT TTCATGGCTC
251 CAAGGACCAT GTATGTGTTG GGGATCTTCA ATATTCATGT TATTTCTCC
301 TTTGGTCTTA TATGATTGTT ACCTTTATGA AGCTTTAGTG ATTACAAAGC
351 ACTTTTTTTG TCCATTTTTC CCTGAGCTTT GTAAACTCTG ATTTGCAGGA
401 TGGCTGGCTG TGGTGAATTT GATCATTCAT TAAACATGCT TCCTACAAAC
451 AGGAAAGCGA ACGAGTCCTG TTCTAATACT GCACCTTCTT TAACCGTCCC
501 TGAATGTGCC ATTTGTCTGC AAACATGTGT TCATCCAGTC AGTCTGCCCT
551 GTAAGCACGT TTTCTGCTAT CTATGTGTAA AAGGAGCTTC ATGGCTTGGG
601 AAGCGGTGTG CTCTTTGTCT ACAAGAAATT CCCGAGGATT TCCTTGACAA
651 GCCCAACCTTG TTGTCACCAG AAGAAGCTCA GGCAGCAAGT AGAGGAAATG
701 GTGAATATGC ATGGTATTAT GAAGGAAGAA ATGGGTGGTG GCAGTACGAT
751 GAGCGCACTA GTAGAGAGCT GGAAGATGCT TTTTCCAAAG GTAAAAAGAA
801 CACTGAAATG TTAATTGCTG GCTTTCTGTA TGTGCTGAT CTTGAAAACA
851 TGGTTCAATA TAGGAGAAAT GAACATGGAC GTCGCAGGAA GATTAAGCGA
901 GATATAATAG ATATACCAA GAAGGGAGTA GCTGGACTTA GGCTAGACTG
951 TGATGCTAAT ACCGTAAACC TAGCAAGAGA GAGCTCTGCT GACGGAGCGG
1001 ACAGTGTATC AGCACAGAGT GGAGCTTCTG TTCAGCCCCT AGTGTCTTCT
1051 GTAAGGCCCC TAACATCAGT AGATGGTCAG TTAACAAGCC CTGCAACACC
1101 ATCCCCTGAT GCAAGCACTT CTCTGGAAGA CTCTTTTGTCT CATTACAAAC
1151 TCAGTGGAGA CAACACAGCT GAAAGGAGTC ATAGGGGAGA AGGAGAAGAA
1201 GATCATGAAT CACCATCTTC AGGCAGGGTA CCAGCACCAG ACACCTCCAT
1251 TGAAGAAACT GAATCAGATG CCAGTAGTGA TAGTGAGGAT GTATCTGCAG
1301 TTGTTGCACA GCACTCCTTG ACCCAACAGA GACTTTTGGT TTCTAATGCA
1351 AACCAGACAG TACCCGATCG ATCAGATCGA TCGGGAAGTG ATCGATCAGT
1401 AGCAGGGGGT GGAACAGTGA GTGTCAAGTG CAGATCTAGA AGGCCTGATG
1451 GACAGTGCAC AGTAACTGAA GTTTAAATAA AAATGTCTTC AGCTCCATGC
1501 TCAAGGTTGA AAGGGTTACC TGTAATTTTC TGCCACATA ACATTATACT
1551 CATCCCTAGT AGTGCAATTT GGGAGTTGGG GTGGGAAGGG GTATGGGAAG
1601 GATAGACTCA TAATTAAAAT GTCTAACATG TCTCTGTTGA GAAATTTATT
1651 TAATGTAAGG AACTTGGGTG TTAATAGTTG AGAGCTGTTT AGTAATAACC
1701 CAGTTTCTCT GAGGTCTGTT TACTTTATAC TTTTAAAAA CTCTGTAGT
1751 TCTTTTGGCC AGTGTGTTTG TATTATCTGT GCATTAATGG TCCTCATCTG
1801 ACTCCTGCAT TGTGTCTTAT TTTTCTGCAT GGATTGGCAT AAGACCATTA
1851 CTAAAATTTG GCACCTGTGA GATGTTTGAT ATTATGAACA GGAACATAA
1901 TTTAATGTAT GAATAGATGT GAATTGGGA TTTCAAAATA GATGAATAAC
1951 AACTATTTTA TAGTAAAGTT ATTGAAATGG AAATGAAAC AGCCAGTAAC
2001 TTATGTTTCA GAATGTTTGT AACACACTTC ATGGTGTTC CATAGGCTTT
2051 GCTGTCTAGT CTTATAGTTT GAGGTTTTTT TGGTCTGCAT TTTTCTTTTT
2101 GATTACAAAA TTTATAATTT AATAAATACT AGAGTTTATC AAAAAAATAA
2151 AAAAAAAG
```

BLAST Results

Entry HSG117 from database EMBL:
human STS SHGC-36270.

Score = 1148, P = 8.9e-45, identities = 240/250

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 400 bp to 1473 bp; peptide length: 358
 Category: similarity to unknown protein
 Prosite motifs: ZINC_FINGER_C3HC4 (51-61)

```

1 MAGCGEIDHS INMLPTNRKA NESCSNTAPS LTVPECAICL QTCVHPVSLP
51 CKHVFCYLCV KGASWLGKRC ALCRQEIPED FLDKPTLLSP EELKAASRGN
101 GEYAWYYEGR NGWWQYDERT SRELEDAFSK GKKNTEMLIA GFLYVADLEN
151 MVQYRRNEHG RRRKIKRDII DIPKKGAVGL RLDCDANTVN LARESSADGA
201 DSVSAQSGAS VQPLVSSVRP LTVSDGQLTS PATPSPDAST SLEDSFAHLQ
251 LSGDNTAERS HRGEGEEDHE SPSSGRVPAP DTSIEETESD ASSDSEDVSA
301 VVAQHSLTQQ RLLVSNANQT VPDRSDRSRG DRSVAGGGTV SVSVRSRRPD
351 GQCTVTEV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27o14, frame 1

TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6,
 N = 2, Score = 165, P = 4.2e-15

SWISSPROT:YWZ6_CAEEL HYPOTHETICAL 39.3 KD PROTEIN C02B8.6 IN CHROMOSOME
 X., N = 2, Score = 136, P = 3.1e-11

>TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6
 Length = 484

HSPs:

Score = 165 (24.8 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 42/106 (39%), Positives = 61/106 (57%)

Query: 75 QEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRN-GWWQYDERTSRELEDAFSKGGK 133
 Q +P LD ++ PEE K Y W Y G+N GWW+++ R RE+E+A++ GK
 Sbjct: 93 QNVPALDLDA-SICDPEERK-----Y-WIYSGKNQGWWRFEPRNEREIEEAYNAGKC 142

Query: 134 NTEMLIAGFLYVADLENMVQYRRNEHRRRKIKR---DIID-IPKKGAVGL 180
 + E++I G YV D +QY R + R +KR D D I KG+AG+
 Sbjct: 143 HCEVVICGRPYVIDFHQFLQYPRGVPNQARHVKRVSADDFDGIGVKGLAGI 193

Score = 96 (14.4 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 19/54 (35%), Positives = 30/54 (55%)

Query: 35 ECAICLQTCVHPVSLP-CKHVFCYLCVKGASW--LGKRCALCRQEIPEDFLDKPT 86
 EC IC + P ++P C H FC++C+KG +G C +CR I + +P+
 Sbjct: 11 ECPICQCKMIVPTTIPACGHKFCFICLKGVMNDMGG-CPMCRGPIDSNIFAQPS 64

Pedant information for DKFZphtes3_27o14, frame 1

Report for DKFZphtes3_27o14.1

[LENGTH] 358
 [MW] 38818.90
 [pI] 5.17
 [HOMOL] TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6 2e-12
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YCR066w] 3e-04
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YCR066w] 3e-04
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR066w] 3e-04

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YCR066w] 3e-04
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 4e-04
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR265w] 4e-04
 [BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
 [PROSITE] MYRISTYL 2
 [PROSITE] AMIDATION 3
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 12
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] ZINC_FINGER_C3HC4 1
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 19.83 %

SEQ MAGCGEIDHSINMLPTNRKANESCSNTAPSLTVPECAICLQTCVHPVSLPCKHVFCYLCV
 SEG
 lrmnd-TTTTTEETTETEEETTETEEHHHHH
 SEQ KGASWLGRKRCALCRQEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRNGWWQYDERT
 SEG
 lrmnd- HHHHHHCCBTTTTCBCGGG-CBCC.....
 SEQ SRELEDAFSKGGKNTTEMLIAGFLYVADLENMVQYRRNEHGRRRKIKRDIIDIPKKGVAGL
 SEG
 lrmnd-
 SEQ RLDCDANTVNLARESSADGADSVSAQSGASVQPLVSSVRPLTSVDGQLTSPATPSPDAST
 SEG
 lrmnd-
 SEQ SLEDSFAHLQLSGDNTAERSHRGEGEDHESPSSGRVPAPDTSIEETESDASSDSEDVSA
 SEG X.....
 lrmnd-
 SEQ VVAQHSILTQORLLVSNANQTVDRSDRSGTDRSVAGGGTVSVSVRSRRPDGQCQTVTEV
 SEG xxx.....
 lrmnd-

Prosites for DKFZphtes3_27o14.1

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	318->322	ASN_GLYCOSYLATION	PDOC00001
PS00004	132->136	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	120->123	PKC_PHOSPHO_SITE	PDOC00005
PS00005	217->220	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	274->277	PKC_PHOSPHO_SITE	PDOC00005
PS00005	325->328	PKC_PHOSPHO_SITE	PDOC00005
PS00005	330->333	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	222->226	CK2_PHOSPHO_SITE	PDOC00006
PS00006	240->244	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	287->291	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00007	98->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00009	66->70	AMIDATION	PDOC00009
PS00009	130->134	AMIDATION	PDOC00009
PS00009	159->163	AMIDATION	PDOC00009
PS00518	51->61	ZINC_FINGER_C3HC4	PDOC00449

Pfam for DKFZphtes3_27014.1

```
HMM_NAME      Zinc finger, C3HC4 type (RING finger)
HMM            *CPICFcTFQlDyPWPfdePmMlPCgHsFCypCIrrW.....CPmC*
               C+IC      L      + P++LPC+H+FCY C++      C +C
Query          36 CAIC-----LQT---CVHPVSLPCKHVFYLCVKGASWLGKRCALC      73
```


DKFZphtes3_28d14

group: testes derived

DKFZphtes3_28d14 encodes a novel 97 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1279 bp

Poly A stretch at pos. 1232, no polyadenylation signal found

```
1 GGAGCTCAGA AGTTGGGCAA AGGTCACAGC AGACTTCCTG AAAAGCAGAC
51 ACTGAGGAAC ACAGTGGAGA GCGGGAGTTC ACAGCGACGC AGCTGAGGAC
101 GACGCAGGAC CTCTCCCAA GGTGCTGCAG CTCCAGACC AGGGGCCAGG
151 GCTGCGGCGA CAGCAGCTCA CCAACCCCTG CTGTGCTCAA GTTCTTGGGG
201 ATTGAGAGCT AAGTTCAAAA TTAGAAACA GTGCCTTAA GACGGGCAAG
251 AAAACCCGGT GTGGGAGTCT GCTCATCTAT GGTTTGTTAC TGCTCTCGCT
301 TTGATATCTT TAAATTCCCTA GGTACCAATG AAAAAGCCAA GTGAACGTGG
351 CAGAGTGAGG AGGAGACAGG AGCGTGTGCA CCTTCCATCT GTGAGAGGCA
401 CACTTCAGTC TGGGTTCAAG ATGCAGAATG GTGCCTACAG CAAAAAAGGCA
451 AAAAACACCC TCCTCCCTTC TTTACCATT GAATGGACAT TTTCTTACC
501 TGTGATCCCA ACAGAAACAG ATCCAGACCT ATCATGTGAA GTCCACGTTT
551 CAGGATCAGA AGTAACCACT TATGGACTG AGCTTACACG GGAAAGTCTA
601 CCCCCGACTC CTTCTGGATA GTAACATACA CAGCTGCATA AAAACGTCTC
651 CAAGGGGACA TACGATGCAT TTGCTTGGTG TCCAGGCCAA GCTCCCCACC
701 GGCGACCTCA CTGTTCCCTTA GAGCTCGAGA GCTCGTCTCC TATCAATCAG
751 AGAACCCCAT CAGCTGTGAC CAACAGAGCT GGAGCCCTCT GTGGAGGGAG
801 CTGACCCAC ACACAGGACA GAGCAGAATC CTGATTATTT TACAACTGC
851 AAACCTTCTG AGTAAGAAGA CAAAAATATA CATTCCAAGG TATCTGTAAA
901 GTGCTTGAAA GATGCAGACA GCTGCACCGA GGGGCTCTGA TCCATCCACA
951 CGCTGCGCTT TGCTGCGGTC ACACACACGG TCTCAGTCAC GTGATGGTTT
1001 TGCTTTTATT TCTTAAACGG CTGAGTGATA ATCCAGCTAG TGTGCAGTCA
1051 TTTTATACCT TTCAATGGGC GTCACCGCAG TGACGCTGCC CCAGCCCCAT
1101 GCTGAGGGCC GACACAATTC ACGGAACAGA TTCATCATAT TTGGTCTTTA
1151 TGTAATAAT AAATGTTTTA AAATTGCCTA AATATAAAAA AAAAAAAAAA
1201 AAAAAAAAAA AAAAAAAAAA AAAGGGCGGC CGAAAAA AAAA
1251 AAAAAAAAAA AAAAAAAAAA GGGCGGCCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 618 bp; peptide length: 97
Category: putative protein

```
1 MKKPSEGRV RRRQERVHLP SVRGTLQSGF KMONGAYSKK KNTLLPSLP
51 FEWTFSLPVI PTETDPLDSC EVHVPGEVLT SLWTELTRES LPPTPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_28d14, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_28d14, frame 1

Report for DKFZphtes3_28d14.1

(LENGTH) 97
 (MW) 10945.56
 (pI) 9.80
 (PROSITE) MYRISTYL 2
 (PROSITE) CAMP_PHOSPHO_SITE 2
 (PROSITE) CK2_PHOSPHO_SITE 2
 (PROSITE) PKC_PHOSPHO_SITE 3
 (KW) All_Alpha
 (KW) LOW_COMPLEXITY 12.37 %

SEQ MKKPSERGRVRRRQERVHLPVSRGTLQSGFKMQNGAYSKKKKNTLLPSLPFEWTFSLPVI
 SEGXX
 PRD cccccchhhhhhhhhhhcc

 SEQ PTETDPDLSCVHVPGSEVTSWLTELTLRESLPPTPSG
 SEG
 PRD cccccccccceccccchhhhhhhhhhhcccccccc

Prosites for DKFZphtes3_28d14.1

PS00004	2->6	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	41->45	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_28d14.1)

DKFZphtes3_2a11

group: testes derived

DKFZphtes3_2a11 encodes a novel 1048 amino acid protein with very weak similarity to mucins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to mucin

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4082 bp

Poly A stretch at pos. 4060, polyadenylation signal at pos. 4034

```
1 GAGGACTGCG AGCACAGCGG CGGCCGGGTG GCGGGGGTGA GTGGGGCCAG
51 CGGGGCTGGA CAGCAGCGGG CCCC GGCGCG CGCCGCCGCG ATCCCTCCCC
101 GCGCCCGCGC AGCACATCGC CGCCGCCGAG ATGGGCCCTC CGCGGCACCC
151 CCAGGCCGGC GAGATAGAAG CGGGCGGTGC GGGCGGCGGG CGGCGGCTAC
201 AGGTGGAAT GAGTTCTCAA CAGTTTCTTC GGTTAGGAGC CCCTTCTACC
251 GGGCTGAGCC AGGCCCTTC TCAGATTGCA AACAGTGGTT CTGCTGGATT
301 GATAAACCCA GCTGCTACAG TCAATGATGA ATCTGGTCGA GATTCTGAAG
351 TCAGTGCCAG GGAGCACATG AGTTCCAGCA GCTCCCTCCA GTCCCGGGAG
401 GAGAAGCAAG AGCCTGTTGT GGTAAGGCCC TATCCACAGG TGCAGATGTT
451 GTCCAGACAC CATGCTGTGC CATCAGCCAC ACCTGTTGCA GTGACAGCCC
501 CGCCAGCACA CCTGACGCCA GCAGTGCCAC TTTCATTTTC GGAGGGACTT
551 ATGAAGCCGC CCCC GAAGCC CACCATGCCT AGCCGTCCTA TTGCTCCTGC
601 TCCACCTTCT ACCCTGTGTC TTCCCCCAA GGTTCAGGG CAGGTTACCG
651 TTACCATGGA GAGTAGCATC CCTCAAGCTT CAGCCATTCC TGTGGCAACA
701 ATCAGTGGAC AACAGGGCCA TCCAGTAAC CTGCATCACA TCATGACTAC
751 AAATGTGCAA ATGTCTATCA TCCGAGCAA TGCTCCTGGG CCCCCTCTTC
801 ACATTGGAGC TTCTCATTTA CCTCGAGGTG CAGCTGCTGC TGCTGTGATG
851 TCCAGTTCTA AAGTAACCAC AGTCCTGAGG CCGACCTCAC AGCTGCCAAA
901 TGCTGCTACT GCTCAGCCAG CAGTACAGCA CATCATTCAC CAACCAATCC
951 AGTCTCGGCC ACCTGTGACC ACCTCCAATG CCATCCCTCC TGCTGTGGTA
1001 GCAACTGTCT CAGCCACCAG AGCTCAGTCT CCACTCATCA CTACGACAGC
1051 GCGCGATGCT ACTGATTGAG CACTTAGTAG GCCAACCTTG TCTATCCAGC
1101 ATCTCTCCATC TGCAGCAATC AGTATTCAGC GTCCTGCCCA GTCACGAGAT
1151 GTCACAACAA GAATCACACT ACCATCTCAC CTGCAATTAG GAGCGCCAAA
1201 ACAGCAGCTT CATACAATGG CTCAGAAAAC AATCTTCAGT ACTGGCACGC
1251 CAGTGGCTGC AGCCACAGTA GCACCTATTT TGGCAACCAA CACCATTCTT
1301 TCAGCGACCA CAGCTGGATC TGTGTACAC ACGCAAGCTC CCACAAGTAC
1351 CATTGTTACC ATGACAGTAC CCTCCCATTC CTCCCATGCT ACTGCTGTGA
1401 CCACCTCAAA CATCCAGTC GCCAAGGTGG TGCCCCAGCA GATCAGGCAC
1451 ACTTCTCTCT GGATCCAGCC AGACTACCTT GCCGAGAGGA GTAGCCTGAT
1501 TCCCATCTCC GGACATCGGG CCTCTCCCAA TCCTGTGGCC ATGGAACCCC
1551 GAAGTGACAA CAGACCGTCT GTTCCCGTTC AGTTCCAATA TTTTTTGCCA
1601 ACTTACCCCT CTCTGCATA CCCACTGGCG GCACATACCT ACACCCCAAT
1651 CACCAATTCC GTGTCCACTA TCCGACAGTA TCCAGTTTCA GCTCAGGCTC
1701 CAAACTCTGC CATCAGAGCT CAGACTGGTG TTGGGGTAGC GTCTACCGTC
1751 CACCTAAACC CCATGCAGTT GATGACAGTG GATGCATCGC ATGCTCGACA
1801 TATTCAGGGG ATCCAGCCAG CACCCATCAG TACCCAGGCT ATCCAGCCGG
1851 CCCCCATTGG GACCCAGGGG ATACAGCCTG CACCATTGG CACACAGGGA
1901 ATTCACCTAG CAACCCCAAT CAACACACAA GGGCTTCAGC CTGCACCTAT
1951 GGGTACTCAG CAGCCTCAGC CTGAAGGAAA GACTTCAGCA GTGGTGTGGG
2001 CAGATGGAGC CACAATTGTG GCCAACCCCTA TTAGCAATCC ATTCAGTGCT
2051 GCTCCAGCAG CAACAACCGT GGTGCAGACC CACAGCCAGA GTGCTAGCAC
2101 CAACGCTCCC GCCCAGGGCT CATCGCCACG GCCAAGCATA CTCCGGAAGA
2151 AACCTGCCAC AGATGGTGCC AAACCCAAAG CTGAAATCCA CGTGTCTATG
2201 GCCACTCCGG TCACTGTGTC CATGGAGACT GTATCCAATC AAAATAATGA
2251 TCAGCCTACC ATTGCCGTCC CTCCAAGTGC CCAGCAGCCC CCACCGACCA
2301 TTCCAACATAT GATTGCAGCA GCCAGTCCCC CGTCACAACC AGCCGTTGCC
2351 CTTTCAACCA TTCTTGAGG GGTCCCCATC ACTCCACCCA TCACCACCAT
2401 TGCACTGCA CCACCTCCAT CAGTCACTGT GGGTGGCAGT CTTTCTCTCG
2451 TCTTGGGCCC TCCCGTTCCT GAAATTAAAG TGAAGAAGA AGTAGAACCA
2501 ATGGATATCA TGAGGCCAGT TTCTGCAGTT CCTCCACTGG CTACCAACAC
2551 TGTGTCTCCA TCTCTTGAT TGCTGGCAAA CAACCTGTCC ATGCTTACAA
2601 GTGACCTACC ACCTGGTGCC TCCCAAGGA AAAAGCCTCG AAAGCAACAG
2651 CATGTGATCT CAACAGAAGA AGGTGACATG ATGGAGACAA ACAGCACTGA
2701 TGATGAGAAG TCCACTGCCA AGAGTCTTCT GGTGAAGGCT GAGAAGCGCA
```



```

2751 AGTCTCCTCC CAAGGAGTAT ATTGATGAGG AAGGTGTGAG ATATGTCCCA
2801 GTGCGTCCAA GACCCCCCAT TACTTTGCTT CGTCACTATC GGAACCCCTG
2851 GAAAGCTGCT TACCACCACT TTCAGAGGTA CAGTGACGTC CGGGTCAAAG
2901 AGGAGAAGAA AGCTATGCTG CAGGAAATAG CTAATCAGAA AGGAGTATCC
2951 TGTCTGTGCTC AAGGCTGGAA AGTCCACCTC TGTGCTGCCC AGTTACTACA
3001 GCTGACGAAT CTAGAACATG ATGTCTATGA AAGACTTACT AACCTGCAGG
3051 AAGGGATTAT CCCAAAGAAA AAAGCAGCAA CAGATGATGA TCTCCACCGA
3101 ATAAACGAAC TGATACAGGG AAATATGCAG AGGTGTAAC TTGTGATGGA
3151 TCAAATCAGT GAAGCCAGAG ACTCCATGCT TAAGGTTTGA GATCATAAAG
3201 ACCGTGCTCT GAAGCTGCTT AACAAGAACG GGACTGTCAA AAAAGTGTCC
3251 AAATTGAAGC GAAAGGAAAA AGTCTAGACC CAGAACAATC AGGAGATTGG
3301 AAGCAAATTT ATGAAGAATG ATGGTGGGGG TGGGGGGAGG GTTTTGGTTT
3351 TTTCCAAAGT GGAACATTGA AATAAAGGAA GTGTTCCCTA GTTCCCGTGT
3401 GAAAGCAGAG GAACCCATGA CATCCAAGGG CGTGAAAGGA TCAGAGCTGA
3451 CTGGACATAG TGAGCTGCCT TCTTGCGTTC GGGTGCACCC CTGTTAAACC
3501 TGATCTGTGT CATAAGTGAC TCCGGATGCA TCAGTGTCCA CCAGTTGGAA
3551 GCAATGACAA GGATGGCTGG CTGGTGTITT TCAGCCTTCC GGTATTAGA
3601 CTGTATTAT CTAGTGGATT CCTGCAGGCC CCATACTGAG CCTGGACTGA
3651 AAGTATCCAC TCGGACCATC TGTATCTCT CTACACTGAA AATAAAACCT
3701 CTTCCACCCA CCCCATTGCG TTCTTCTGCC TGACCTTCAA ATGCCATGT
3751 TGGCCTTTTA CAGCAGTGCC ACGGCACCAA GCGAGCTGCC ACATCTCACA
3801 CTCTAAAGGG TTTGAATAT TAGTTCTTGT CATTTTTTAA AAAAAACCT
3851 TCCCAAGTGA AATTGTTATA TCGTCTGTCT TCGTGTGTC AGAAGTGGGT
3901 TTTTGTGGAG GTTCAGAGCA GGCAACACCA TAAGTTGCTC TCAGATCCTT
3951 GTTCTGAAGT ACATTCTTGG TTATCTGTAC TTCTGTAGCT GGTGTGATGC
4001 TGTAAATTGT ATGTACCACA CATCTCCAGA CGTTAATAAA GGACTCAAAG
4051 AGGTTTTTGT AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 131 bp to 3274 bp; peptide length: 1048
 Category: similarity to known protein

```

1 MGPPRHPQAG EIEAGGAGGG RRLQVEMSSQ QFRLGAPST GLSQAPSQIA
51 NSGSAGLINP AATVNDESGR DSEVSAREHM SSSSSLSQRE EKQEPVVVRP
101 YPQVQMLSTH HAVASATPVA VTAPPAHLTP AVPLSFSEGL MKPPPKPTMP
151 SRPIAPAPPS TSLSPKVPVG QVTVMESSI PQASAI PVAT ISGQOQHPSN
201 LHHIMTTNVQ MSIIRSNAPG PPLHIGASHL PRGAAAAAVM SSSKVTTVLR
251 PTSQLPNAAT AQPAPVQHIIH QPIQSRPPVT TSNAIIPAVV ATVSATRAQS
301 FVITTTAAHA TDSALSRPTL SIQHPPSAAI SIQRPAQSRD VTTRITLPSH
351 PALGTPKQQL HTMAQKTIFS TGTEVAAATV APILATNTIP SATTAGSVSH
401 TQAPTSTIVT MTVPSSHSSHA TAVTTSNIPV AKVVPQIITH TSPRIQPDYP
451 AERSSLIPIG GHRASPNPVA METRSDNRPS VPVQFYFLP TYPPSAYPLA
501 AHYTYPTISS VSTIRQYPVS AQAPNSAITA QTGVGVASTV HLNPMQLMTV
551 DASHARHIQ IQPAPISTQG IQPAPIGTPG IQPAPLGTQG IHSATPINTQ
601 GLQPAPMGTO QPQPEGKTSV VVLADGATIV ANPISNPFSV APAATTVVQT
651 HSQSASTNAP AQGSSPRPSI LRKKPATDGA KPKSEIHVSM ATPVTVSMET
701 VSNQNDQFT IAVPPTAQQP PPTIPTMIAA ASPPSQPAVA LSTIPGAVPI
751 TPIITTIAAA PPPSVTVGGS LSSVLGPPVP EIKVKEEVEP MDIMRPVSAV
801 PPLATNTVSP SLALLANNLS MPTSDLPPGA SPRKKPRKQQ HVISTEEGDM
851 METNSTDDEK STAKSLLVKA EKRKSPPKEY IDEEGVRYVP VRPRPPITLL
901 RHYRNPWKAA YHHFQRYSDV RVKEEKKAML QEIANQKQVS CRAQGWKVHL
951 CAAQLLQLTN LEHDVYERLT NLQEGIIPKK KAATDDDLHR INELIQGNMQ
1001 RCKLVMDQIS EARDSMLKVL DHKDRVLKLL NNGTGVKKVS KLKRKEKV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2all, frame 2

SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2) ., N = 1,
 Score = 334, P = 2.4e-25

PIR:A43932 mucin 2 precursor, intestinal - human (fragments), N = 1,
Score = 321, P = 3.2e-24

TREMBL:D88440_1 product: "high molecular mass nuclear antigen"; Gallus
gallus mRNA for high molecular mass nuclear antigen, partial cds., N =
1, Score = 312, P = 8.3e-24

PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast
(Saccharomyces cerevisiae), N = 1, Score = 300, P = 2.1e-22

>SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
Length = 5,179

HSPs:

Score = 334 (50.1 bits), Expect = 2.4e-25, P = 2.4e-25
Identities = 184/770 (23%), Positives = 263/770 (34%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3471 VTPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPT 3530

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3531 TTPITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQT-PTTTPITTTTITVPTPT 3589

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3590 PTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTPI 3649

Query: 269 IHQIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3650 TTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPT 3706

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3707 PTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTPIIT 3766

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSHATAVTTSNIPVAKVVPQQIHTTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3767 TTTVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTPTGT 3825

Query: 444 RIQPDYPAERSSLIPIGHRASPNFVAMETRSNRPSPVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3826 QTPTTTTITTTTITVPTPTPTPTGTQTPTTTTITTTTITVPTPTPTPTGT--TQTP 3874

Query: 503 TYTPTTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 3875 TTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTPTGTQTPTTTTITTTTITVPT 3932

Query: 561 IQPAPISTQGIQAPIGTPIGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 3933 PTPTPTGTQTPTTTTITTTTITVPTPTPTGTQT-TPTTTITTTTITVPTPTPTGTQTPTT 3991

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P+ T T T +Q+ +T ++ P+
Sbjct: 3992 TPITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPT 4051

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4052 PTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTPIIT 4111

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPS-----VTVGSLSSVLGP-PVPEI 782
P+ T P PIT TT+ P P+ T + ++ + P P P
Sbjct: 4112 TTTVPTPTPTGTQT-PTTTPITTTT-TTVPTPTPTGTQTPTTTTITTTTITVPTPTPTG 4169

Query: 783 KVKEEVEPMDIMRPVSAVP-PLATNTVSPSLALLANLMSMPTSDLPFGASPRKKPRKQKH 841
P+ V+ P P T T P+ A + TS+ PP +S + R
Sbjct: 4170 TQTPTTTTITTTTITVPTPTPTGTQTGPPTHTSTAPIAELTTSNPPESSTPQTSRSTSS 4229

Query: 842 VISTEEGDMMET 853
+ TE ++ T
Sbjct: 4230 PL-TESTLLST 4240

Score = 328 (49.2 bits), Expect = 1.0e-24, P = 1.0e-24
Identities = 180/745 (24%), Positives = 254/745 (34%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3540 VTPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPTTTTITTTTITVPTPTPTGTQTPT 3590

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3600 TTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQT-PTTTPITTTTTVTPT 3658

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3659 PTPGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPIT 3718

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3719 TTTTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVT---PTPT 3775

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAOKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3776 PTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITT 3835

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3836 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTPTGT 3894

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3895 QTPTTTPITTTTTVT-----PTPTPTGTQTPT-----TTPITTTTTVTPTPTPTG--TQTP 3943

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT TV T Q T
Sbjct: 3944 TTTPTTTTTVTPTPTGTQTPTTTPITTTTTVTPTPT--TGTQTPTTTPITTTTTVT 4001

Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 4002 PTPPTGTQTPTTTPITTTTTVTPTPTGTQT-TPTTTPITTTTTVTPTPTGTQTPTT 4060

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 4061 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPT 4120

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4121 PTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITT 4180

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAA-PPPSVTVGSLSSVLGPPVPEIKVKEE 787
P+ T P T PI + + PPP + + S P +
Sbjct: 4181 TTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPFESSTPQTSRSTSSPLTESTLLST 4240

Query: 788 VEPMDIMRPVSAVPPPLATNTVSPSLALLANNLSMP--TSDLPFGASPR 833
+ P M S PP +T T +P+ + LS P T+ PPG R
Sbjct: 4241 LPPAIE--TSTAPP-STPT-APTTTSGGHTLSPPSSTTSPPGTPTR 4284

Score = 325 (48.8 bits), Expect = 2.2e-24, P = 2.2e-24
Identities = 186/782 (23%), Positives = 261/782 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3494 VTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPT 3553

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3554 TTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQT-PTTTPITTTTTVTPT 3612

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3613 PTPGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPIT 3672

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3673 TTTTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVT---PTPT 3729

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAOKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3730 PTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITT 3789

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3790 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTGTQTPTTTPITTTTTVTPTPTPTGT 3848

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3849 QTPTTTPITTTTTVT-----PTPTPTGTQTPT-----TTPITTTTTVTPTPTPTG--TQTP 3897

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560

T TPIT++ + T P Q P + IT TV T Q T
 Sbjct: 3898 TTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT--TGTQTPTTPIITTTTPT 3955
 Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 3956 PTPTPTGTQTPTTTPIITTTTPTPTPTGTQ-TPTTTPIITTTTPTPTPTGTQTPTT 4014
 Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 4015 TPIITTTTPTPTPTGTQTPTTPIITTTTPTPTPTGTQTPTTPIITTTTPTPTPT 4074
 Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P
 Sbjct: 4075 PTGTQTPTTPIITTTTPTPTPTGTQTPTTPIITTTTPTPTPTGTQTPTTPIIT 4134
 Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P I V
 Sbjct: 4135 TTTVPTPTPTGTQT-PTTPIIT---TTTTVPTPTPT--GTQT---PTTPIITTTT 4184
 Query: 789 EPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPAGSPRKKPRKQHVISTEEG 848
 P PP T+T +P L +N PS P + P + + +
 Sbjct: 4185 TPTPTGTGTGPPTHST-APIAELTTSN-PPESSTPQTSRSTSSPLTESTLLSLTP 4242
 Query: 849 DMMETNSTDDEKSTAKSLLVKAERKSP 877
 +E ST + SPP
 Sbjct: 4243 PAIEMTSTAPPSTPTAPTTSGGHTLSPP 4271
 Score = 324 (48.6 bits), Expect = 2.8e-24, P = 2.8e-24
 Identities = 170/717 (23%), Positives = 248/717 (34%)
 Query: 95 PVVVRPYQVQMLSTHHAVASATP--VAVTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSR 152
 P P P +T + +P T PP TP+ P++ + + P P+ P
 Sbjct: 1401 PPTTTPSPPTTTTLPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPL-PTTTPSPPIIS 1459
 Query: 153 PIAPAPPSTLSLPPKVPQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 PP+T PP T S + P T + P I +
 Sbjct: 1460 TTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPS---PPMTTPTTPASTTT 1516
 Query: 213 IIRSNAPGPHLHIGASHLPRGAAAAVMSSSKVTTVLRPTSQ--LPNAATAQPAVQHIIH 270
 + + P PP + P S T + PTS LP T P
 Sbjct: 1517 LPPTTTPSPPTTTTTPPP-----TTTSPPTTTPITPPTSTTTLPTTTPSPPTTTT 1571
 Query: 271 QPIQSRP-PVTTSNAIIPAVVATVSA-TRAQSPVITTTAAHATDSALSRLTSLIQHPESA 328
 P + P P TT+ PP + T T SP TTT + S PT + PP++
 Sbjct: 1572 PPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1631
 Query: 329 AISIQRAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIESTGTPVAAATVAPILATNT 388
 ++ T T P TP T I +T TP T + + T
 Sbjct: 1632 TTTLPPTTTPSPPTTTTTP--PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSSPITTP 1689
 Query: 389 IPSATTAGSVSHTQAPTSTIVTMTVPSHSSHATAV-TTSNIPVAKVVPQITHTSPRIQP 447
 P TT + ST P+S I T T PS ++ + TT P P T T + P
 Sbjct: 1690 SPPTTMTTPSPPTTTPSPPTTTPSSTTTPSPPTTMTTPSPPTTTPSPPTTMTTLP 1749
 Query: 448 DYPAERSSSLIPISGHRASPNPVMETRSNRPVVPV-QFYFLPTYPSPAY-P-----LA 500
 + + P+ P T + P VP+ + +L + P+ + P L
 Sbjct: 1750 TTTSSPLTTTLPSPITPPTFSPTTTPPTPCVPLCNWTGWLDGKPNFHKPGGDELI 1809
 Query: 501 AHTYTPITSSVSTIR--QYP-VSAQAPNSAITAQTVGV-VASTVHLNPMQLMTVDASHAR 556
 P ++ + R YP V + VG + P ++ + A
 Sbjct: 1810 GDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPM-AFCLN 1868
 Query: 557 HIQGIQAPISTQGIQAPIGTPIGQ-PAPLGTQGIHSATPINTQGLQAPMGTQQPQ-- 613
 + +Q TQ P + T + P P T I + T + P P GTQ P
 Sbjct: 1869 YEINVQCCECVTQ---PTTMTTTTENPTPTTPIITTTTPTPT--PTPTGTQTPTT 1922
 Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 1923 PITTTTPTPTPTPTGTQTPTTPIITTTTPTPTPTGTQTPTTPIITTTTPTPTPT 1982
 Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMIA 729
 T P + TP +T + T P PT Q P T P
 Sbjct: 1983 TGTQTPTTPIITTTTPTPTPTGTQTPTTPIITTTTPTPTPTGTQTPTTPIIT 2042
 Query: 730 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEVE 789
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 2043 TTVPTPTPTGTQT-PTTPIIT---TTTTVPTPTPT--GTQTPTTPIITTTTPTPTPT 2096
 Query: 790 PMDIMRPVSAVPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 2097 PTGTQTPTT-PIITTTTPTPT 2117

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2068 VTPTPTPTGTQTPTTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPT 2127

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2128 TTPITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQT-PTTTPITTTTIVTPT 2186

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2187 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2246

Query: 269 IHQPIQSRPPVTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLTSLIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2247 TTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2303

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2304 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2363

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2364 TTTVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTPTGT 2422

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2423 QTPTTTPIITTTTIVT-----PTPTPTGTQTPT-----TTPITTTTIVTPTPTGT--TQTP 2471

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2472 TTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTPT--TGTQTPTTTPIITTTTIVT 2529

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL--QPAPMGTOQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2530 PTPTPTGTQTPTTTPIITTTTIVTPTPTGTQ-TPTTPIITTTTIVTPTPTGTQTPTTT 2588

Query: 614 -PEGKTSAAVVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2589 TPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPT 2648

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 2649 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2708

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTIAAPPSPSVTVGGSLSVLPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 2709 TTTVTPTPTPTGTQT-PTTPIIT---TTTTVTPTPTPT--GTQTPTTTPIITTTTIVTPTPT 2762

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811
P + P T TV+P+
Sbjct: 2763 TPTGTQTPTTT-PIITTTTIVTPT 2784

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2206 VTPTPTPTGTQTPTTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPT 2265

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2266 TTPITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQT-PTTTPITTTTIVTPT 2324

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2325 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2384

Query: 269 IHQPIQSRPPVTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLTSLIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2385 TTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2441

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2442 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2501

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2442 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2501

Sbjct: 2502 TTTVTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPT-TPTGT 2560

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+

Sbjct: 2561 QTPTTTTPITTTTVT-----PTPTPTGTQTPT-----TTPITTTTTTPTPTPTG--TQTP 2609

Query: 503 TYTPTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
 T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2610 TTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPT--TGTQTPTTTTPITTTTTPT 2667

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQOPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2668 PTPTPTGTQTPTTTTPITTTTTPTPTPTGTQ-TPTTTTPITTTTTPTPTPTGTQTPTT 2726

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2727 TPITTTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPT 2786

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P

Sbjct: 2787 PTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPIIT 2846

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2847 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTTPITTTTTPTPT 2900

Query: 789 EPMDIRPVSAPVPLATNTVSPS 811
 P P + P T TV+P+

Sbjct: 2901 TPTGTQTPTTT-PITTTTTTPTPT 2922

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPVQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T ++ T V T P TP + + P P PT P

Sbjct: 2321 VTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPT 2380

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHFNSLNHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 2381 TPIITTTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQT-PTTTPITTTTTTPTPT 2439

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I

Sbjct: 2440 PTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTPI 2499

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P

Sbjct: 2500 TTTTPTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPT 2556

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI

Sbjct: 2557 PTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPIIT 2616

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQOITHTSP 443
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T

Sbjct: 2617 TTTVTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTPTGT 2675

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+

Sbjct: 2676 QTPTTTTPITTTTVT-----PTPTPTGTQTPT-----TTPITTTTTTPTPTPTG--TQTP 2724

Query: 503 TYTPTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
 T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2725 TTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPT--TGTQTPTTTTPITTTTTPT 2782

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQOPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2783 PTPTPTGTQTPTTTTPITTTTTPTPTPTGTQ-TPTTTTPITTTTTPTPTPTGTQTPTT 2841

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2842 TPITTTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPT 2901

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P

Sbjct: 2902 PTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPITTTTTPTPTPTGTQTPTTTTPIIT 2961

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2962 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTTPITTTTTPTPT 3015

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 3016 TPTGTQTPPTT-PITTTTIVTPT 3037

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + TV T P TP + + P P PT P
 Sbjct: 2390 VTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTP 2449

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 2450 TTPITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPT 2508

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 2509 PTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPI 2568

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTAHATDSALSRLTSLIQHPPSA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 2569 TTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVT---PTPT 2625

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 2626 PTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIIT 2685

Query: 386 TNIT-PSATTAGSVSHTQAPTSTIVTMT-VPSHSHATAVTTSNIPVAKVVPQQIHTSP 443
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
 Sbjct: 2686 TTVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGT 2744

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 2745 QTPTTTPITTTTIVT---PTPTPTGTQTP---TTPITTTTIVTPTPTPTG--TQTP 2793

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARIHQ 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 2794 TTPITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTG--TGTQTPPTTPIITTTTIVT 2851

Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 2852 PTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQ-TPITTPITTTTIVTPTPTPTGTQTPPT 2910

Query: 614 -PEGKTSAAVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 2911 TPITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPT 2970

Query: 672 RKKPATDGAKEPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P
 Sbjct: 2971 PTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIIT 3030

Query: 729 AAASPPSQPAVALSTIPGAVPITPFIITIAAAPPSVTVGGSLSVLPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 3031 TTTTIVTPTPTPTGTQTPPTTPIIT---TTTTVPTPTPT--GTQTPPTTPIITTTTIVTPTPT 3084

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 3085 TPTGTQTPPTT-PITTTTIVTPT 3106

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + TV T P TP + + P P PT P
 Sbjct: 2459 VTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTP 2518

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 2519 TTPITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPT 2577

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 2578 PTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPI 2637

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTAHATDSALSRLTSLIQHPPSA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 2638 TTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVT---PTPT 2694

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 2695 PTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIITTTTIVTPTPTPTGTQTPPTTPIIT 2754

769

Sbjct: 3169 TTTVTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTTTTPTPT 3222
 Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 3223 TPTGTQTPTTT-PITTTTTVTPT 3244

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3080 VTPTPTPTGTGTPTTTTPTTTTPTPTPTGTGTPTTTTPTTTTPTPTPTGTGTPT 3139
 Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 3140 TTPITTTTPTPTPTGTGTPTTTTPTTTTPTPTPTGTGT-PTTPTITTTTPTPT 3198
 Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 3199 PTPTGTQTPTTTPTTTTPTPTPTGTGTPTTTPTTTTPTPTPTGTQTPTTTPI 3258
 Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 3259 TTTTPTPTPTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTPTPT 3315
 Query: 329 AISIQRPQASRDVTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAA--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 3316 PTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTT 3375
 Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHSP 443
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
 Sbjct: 3376 TTTVTPTPTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTPTPTGT 3434
 Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVFPVQFQYEL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 3435 QPTTTPTTTTPT 3483
 Query: 503 TYTPITSSVS-TIRQYPVSAQAQNSA-ITAQTGCVGVASTVHLNPMQLMTVDASHARHIQ 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 3484 TTTPTTTTPTPTPTPTGTQTPTTTPTTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3541
 Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 3542 PTPTPTGTQTPTTTPTTTTPTPTPTPTGTQT-PTTTPTTTTPTPTPTPTGTQTPTT 3600
 Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 3601 TPTTTTPTPTPTPTGTQTPTTTPTTTTPTPTPTPTGTQTPTTTPTTTTPTPTPTPT 3660
 Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P
 Sbjct: 3661 PTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTPTGTQTPTTTPTTT 3720
 Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPSVTVGGSLSVLPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 3721 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTTTTPTPTPT 3774
 Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 3775 TPTGTQTPTTT-PITTTTTVTPT 3796

Score = 313 (47.0 bits), Expect = 4.2e-23, P = 4.2e-23
 Identities = 169/695 (24%), Positives = 245/695 (35%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3655 VTPTPTPTGTGTPTTTPTTTTPTPTPTGTGTPTTTPTTTTPTPTPTGTGTPT 3714
 Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 3715 TTPITTTTPTPTPTGTGTPTTTPTTTTPTPTPTGTGT-PTTPTITTTTPTPT 3773
 Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 3774 PTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPI 3833
 Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 3834 TTTTPTPTPTPTGTQTPTTTPTTTTPTPTPTGTQTPTTTPTTTTPTPTPTPTPT 3890

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 3891 PTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIIT 3950

Query: 386 TNTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHSTP 443
 T T+ P+ T G+ + T P+ T T+T P+ + T TT V P T T
 Sbjct: 3951 TTTVTPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTP-TPTGT 4009

Query: 444 RIQPDYPAERSSSLIPISGHRASPNPVMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 4010 QTPTTTTPIITTTTTP-----PTPTPTGTQTP-----TTPITTTTTPPTPTGT--TQTP 4058

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 4059 TTPITTTTTPPTPTPTGTQTPTTTPIITTTTTPPTPTPT--TGTQTPTTTPIITTTTTP 4116

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTOGL---QPAPMGTOQPQP 614
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 4117 PTPTPTGTQTPTTTPIITTTTTPPTPTGTQ-TPTTPIITTTTTPPTPTGTQTP-T 4174

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRK 674
 T+ + T+ P P T ++ ++N P+ S+P+ S
 Sbjct: 4175 ---TTPITTT---TTVTPTPTGTQTPPTHTSTAPIAELTTSNPPESSTPQTSRSTSS 4229

Query: 675 PATDGAKPKSEIH--VSMATPVTVSMETVSNQNDQPTIAVPP-TAQQPP--PTIPTMIA 729
 P T+ S + + M + S T + T++ PP T PP PT T
 Sbjct: 4230 PLTESTTLLSLTPAIEMTSTAPPSTPTAPTTTSGGHTLSPPSTTSPPGTPTRGTTG 4289

Query: 730 AASPPSQPAVALSTI----PGAVPITPP--ITTIAAAP-PPSVTVGGSLSSVLGPPVPEI 782
 ++S P+ V +T P P++ P IT P P SV + L+ P E+
 Sbjct: 4290 SSSAPTSTVQTTTTSAWTPPTPLSTPFIIRTGLRYPSSVLICCVLNDYYAPGEEV 4349

Score = 279 (41.9 bits), Expect = 1.8e-19, P = 1.8e-19
 Identities = 138/540 (25%), Positives = 194/540 (35%)

Query: 278 PVTTSSNAIPPAVVATVSATRAQSPVITTTAAH-----ATDSALSRP--TLSIQHPPSAA 329
 P+TT+ + P T + T +P+ TTT T + + P T + P
 Sbjct: 1946 PTTTTTTPPTPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPT 2005

Query: 330 ISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILAT 386
 Q P + TT P+ GT + T + T TP T PI T
 Sbjct: 2006 TGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTT 2065

Query: 387 NTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHSTSPR 444
 T+ P+ T G+ + T P+ T T+T P+ + T TT V P T T +
 Sbjct: 2066 TTVTPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTP-TPTGTQ 2124

Query: 445 IQPDYPAERSSSLIPISGHRASPNPVMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAHT 503
 P ++ + +P P +T + + P+ + PT P+ T
 Sbjct: 2125 TPTTTTPIITTTTTP-----PTPTPTGTQTP-----TTPITTTTTPPTPTGT--TQTP 2173

Query: 504 YTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGI 561
 TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 2174 TTPITTTTTPPTPTPTGTQTPTTTPIITTTTTPPTPTPT--TGTQTPTTTPIITTTTTP 2231

Query: 562 QPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTOGL---QPAPMGTOQPQP-- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 2232 TPTPTGTQTPTTTPIITTTTTPPTPTGTQ-TPTTPIITTTTTPPTPTGTQTPTTT 2290

Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 2291 PTTTTTTPPTPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPT 2350

Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMIA 729
 T P + TP +T + T P PT Q P T P
 Sbjct: 2351 TGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTT 2410

Query: 730 AASPPSQPAVALSTIPGAVPITPPITTTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVE 789
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 2411 TTVTPTPTGTQTP-TTTPIT---TTTTTTPPTPT--GTQTPTTTPIITTTTTPPT 2464

Query: 790 PMDIMRPVSAPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 2465 PTGTQTPTTT-PIITTTTTPPT 2485

Score = 265 (39.8 bits), Expect = 5.8e-18, P = 5.8e-18
 Identities = 179/746 (23%), Positives = 257/746 (34%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKTTPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3678 VTPTPTPTGTQTPTTTPIITTTTTPPTPTGTQTPTTTPIITTTTTPPTPTGTQTP 3737

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 3738 TTPITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQT-PTTPTITTTTTVTPT 3796

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRLPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 3797 PTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTI 3856

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLTSLIQHPSPA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 3857 TTTTPTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVT---PTPT 3913

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 3914 PTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITT 3973

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTTSP 443
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
 Sbjct: 3974 TTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGT 4032

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 4033 QTPTTPTITTTTTVT-----PTPTPTGTQTPT----TPTITTTTTVTPTPTPTG--TQTP 4081

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 4082 TTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPT--TGTQTPTTPTITTTTTVT 4139

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTTQQPQP 614
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 4140 PTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQ-TPTTPTITTTTTVTPTPTPTGTQTGPP 4198

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPA---ATTVVQTHSQSA-STNAPA--QGSSPRP 668
 TS +A+ T +NP P S+ P +T+ T S + ST PA S+ P
 Sbjct: 4199 T-HTSTAPIAELTT--SNP--PPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPP 4253

Query: 669 SILRKKPATDGA KPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQFPPTIPTMI 728
 S T G S + +P + ++ PT + T T PT
 Sbjct: 4254 STPTAPTTSGGHTLSPPTTTSPPGTPTRGTTTGSSAPTPTSTVQTTTSAWT-PTPT 4312

Query: 729 AAASPPSQPAVALSTIPGAVPITPPTITIAAAPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
 ++P L P +V I + AP V G+ + E
 Sbjct: 4313 PLSTPSIIRTGLRYPSSVLICCVLNDTYAPGEEV-YNGTYGDTCYFVNCSLSCTLEF 4371

Query: 789 EPMIDMRPVSAVPPLATNTVSPSLALLANLMSPTSDLPFGASPRKKPRKQOH 841
 S P + +T +PS ++ S PT P P P +Q++
 Sbjct: 4372 YNWSCPSTPSPTPTPSKSTPTPSKP--SSTPSKPTPGTKPPECDFDPPRQEN 4422

Score = 254 (38.1 bits), Expect = 8.7e-17, P = 8.7e-17
 Identities = 167/697 (23%), Positives = 245/697 (35%)

Query: 115 SATPVAVTAPPAHLTPAVPLSFSEGLMKPPPK--PTMPSP-PIAPAPPSTLSLPPKV-PG 170
 S + T PP TP+ P + + PPP P+ P+ PI P P ST +LPP P
 Sbjct: 1587 SPPTITTTTTPTPTTPTSPPTTTT---TPTPTTPTSPPTTPTITP-PTSTTTLPPTTTPS 1642

Query: 171 QVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHL 230
 T + P + P T + + TT I + P PP +
 Sbjct: 1643 PPPTTTTTPTPTTPTSPPTTPTTPTPTTPTTPTTPTTPTSSPI--TTTSPPTTMTTTPS 1700

Query: 231 PRGAAAAVMSSSKVTVLRLPTSQLPNAATAQPAVQHIIHQPIQS-RPPVTTSSNAIPPAV 289
 P SS +TT P+S + P P + PP TT +PP
 Sbjct: 1701 P-----TTTSSPITTTTTPTSS---TTTSPPTTMTTPTPTTPTPTTPTTMTLPTT 1751

Query: 290 VATVSATRAQSPVIIT-TAAHATDSALSRLTSLIQH----PPSAAISIQRPAQSRDVTTR 344
 ++ T P IT T + + + + P + + + S + +P ++
 Sbjct: 1752 TSSPLTTTLPSPITPTPTSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDELIGD 1811

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATN-----TIPSATTAGS 397
 + P A + + ++ I G V ++ N IP A
 Sbjct: 1812 VCGPGWAANISCRATMYP--DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNY 1869

Query: 398 VSHTQAPTSTI--VTMTVPSHSSHATAVTTSNIPVAKVVPQQIHTTSPRIQPDYPAERSS 455
 + Q TMT + + + T TT+ I V T T + P ++
 Sbjct: 1870 EINVCQCECVTQPTTMTTIT-TENPTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTT 1928

Query: 456 LIPISGHRASPNPVMETRSNRPSPVPVQFQYFL-PTYPPSAYPLAAHTYPTITSSVS-T 513
 + +P P +T + + P+ + PT P+ T TPIT++ + T
 Sbjct: 1929 TVT-----PTPTPTGTQTPT----TPTITTTTTVTPTPTPTG--TQPTTPTITTTTTVT 1977

Query: 514 IRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQIQPAPISTQGIQ 572

Sbjct: 1978 P TPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2035
 Query: 573 PAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQQPQ--PEGKTSAVVLA 624
 PI T P P GTQ + TPI T P P GTQ P P T+ V
 Sbjct: 2036 TTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2094
 Query: 625 DGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILRKKPATOGAKPK 683
 T P + P + T T T +Q+ +T ++ P+ T P
 Sbjct: 2095 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTP 2154
 Query: 684 SEIHVS MATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMIAAASPPSQPAVA 740
 + TP +T + T P PT Q P T P P+
 Sbjct: 2155 TTTTPTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2214
 Query: 741 LSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEVEPMDIMRPVSAV 800
 T P PIT TT P P+ T G+ + P V P P +
 Sbjct: 2215 TQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTPTPTGTQTPTT- 2267
 Query: 801 PPLATNTVSPS 811
 P T TV+P+
 Sbjct: 2268 PITTTTTVTPT 2278

Score = 243 (36.5 bits), Expect = 1.3e-15, P = 1.3e-15
 Identities = 110/406 (27%), Positives = 154/406 (37%)

Query: 121 VTAP-PAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPQVTVTMES 179
 +T P P TP+ P + + L P P+ P+ PP+T PP T + ++
 Sbjct: 1396 ITTSPPTTTPSPPTTTTTL-PPTTTPSPPTTTTTTPPTTTPSPPTT--TTPLPTT 1452
 Query: 180 IPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPHIGASHLPRGAAAAAV 239
 P P++T + P+ TT + P PP + P
 Sbjct: 1453 TPSP---PISTTTTP--PTTTPSPPTTTPSP--TTTSPPTTTTTTPPP-----TT 1498
 Query: 240 MSSSKVITVLRP---TSQLPNAATAQPAVQHIHQPIQSRP-PVTTSNAI PPAVVATVSA 295
 S +TT + P T+ LP T P P + P P TT+ PP T+
 Sbjct: 1499 TPSPPTTTPITPPASTTTLPPTTTPSPPTTTTTTPPTTTPSPPTTTPITPPTSTTTLP 1558
 Query: 296 TRAQSPVITTTAAHATDSALSRPTLSIQHPSSAAISIQRPASRDV-TTRITLPSHPALG 354
 T SP TTT + S PT + PP+ + P + TT T P P
 Sbjct: 1559 TTTSPPTTTTTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSPPTTTTTTP--PPTT 1616
 Query: 355 TPKQQLHTMAQKTIFSTGTPVAAATVAPILATNTIPSATTAGSVSHQAPTSTIVMTVP 414
 TP T +T P T+P T T P TT S T P+ I T T P
 Sbjct: 1617 TPSPPTTTPITPPTSTTTLP-PTTTPSPPTTTTTTPPTTTPSPPTTTPSPPTTTTTTP 1675
 Query: 415 SHSSHATA-VTTSNIPVAKVVPQQITHTSPIQPDYPAERSSSLIPISGHRASPNVAMET 473
 ++ ++ +TT+ P + T SP P P ++ P S SP P M T
 Sbjct: 1676 PPTTTPSSPITTTTPSPPTTMM---TTPSPPTTTPSSPITTTT-PSSTTTPSPPTTMTT 1730
 Query: 474 RSDNR-PSVFPVQFYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNS 526
 S PS P LP S+ PL T TP+ S++ P S P +
 Sbjct: 1731 PSPTTTPSPPTTMTTLPPTTSS-PL---TTPLPSPITPPTSPFSTTTPTT 1780

Score = 189 (28.4 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 92/374 (24%), Positives = 133/374 (35%)

Query: 439 THTSPRIQPDYPAERSSSLIPISGHRASPNVAMETRSDNRSPVVPVQFYF-LPTYPPSAY 497
 T + P P P ++ +P + + P PS P+ LPT PS
 Sbjct: 1398 TPSPPTTTPSPPTTTTTLPPTTTPSPPTTTTTTPPTTTPSPPTTTTTPLPTTTPSP- 1456
 Query: 498 PLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQL-MTVDASHAR 556
 P++ T P T++ S P S T T +T PM +T AS
 Sbjct: 1457 PISTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTTTTPPTTTPSPPTTTPITPPASTTT 1516
 Query: 557 HIQGIQAPISTQGIQAPIGTPGIQAPLGTQGIHSATPINTQGLQAPMGTOQQPQEG 616
 P+P +T P P TP +P T I P +T L P T P P
 Sbjct: 1517 LPPTTTPSPPTTTTTTPPTTTP---SPPTTTP--TPPTSTTTLP--TTTSPPPP 1566
 Query: 617 KTSVVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP--AQGSSPRPSILRKK 674
 T+ T +P P + P+ T+ T +T +P ++P P+
 Sbjct: 1567 TTTTT---PPPTTTPSP---PTTTPSPPTITTTTPPTTTPSPPTTTTTTPPTTTPSP 1620
 Query: 675 PATDGAKPKSEIHVS MATPVTVSMETVSNQNDQPTIAV-PPTAQPPPTIPTMIAA--A 731
 P T P + + P T + PT PPT P P I T
 Sbjct: 1621 PTTTPTPTPS--TTTLPTTTPSPPTTTTTTPPTTTPSPPTTTTTTPSPPTTTTTTPPT 1678
 Query: 732 SPFSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPV-----PEIKVK 785
 + PS P + P TP TT ++P + T S ++ PP P
 Sbjct: 1679 TTPSSPITTTTPSPPTTMTTTPSPPTTTPSSPITTTTTTPSSTTTPSPPTTMTTTPSPPTTTPS 1738

Query: 786 EEVEPMDIMRPVSAVPPLATNTVSPSL 812
 M + P + PL T + PS +
 Sbjct: 1739 PPTTTMTTLPPTTSSPLTTTLPPLSI 1765

Score = 185 (27.8 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 71/270 (26%), Positives = 99/270 (36%)

Query: 563 PAPISTQGIQAPIGTPGIQAPPLGTQGIHSATP---INTQGLQAPMGTQQPQ---PEG 616
 P+P +T P P TP P T + + TP I+T P P T P P
 Sbjct: 1422 PSPPTTTTTPPTTTPS-PPITTTTLPPTTTPSPPISTT-TTPPTTTPSPPTTTPSP 1479

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPA 676
 T+ T P + P +P TT + T S +T P SP + P
 Sbjct: 1480 PTTTSPPTTTTTPPTTTP---SPPMTTPI-TTPASTTTLPTTTPSPPTTTTTPPP 1535

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQ 736
 T P + TP+T T + P+ P T PPPT + PS
 Sbjct: 1536 TTTSPPT-----TTPITPPTSTTTLPPTTTPS-PPPTTTTTPPTTTPSPPTTTPSP 1588

Query: 737 PAVALSTIPGAVPITPPTTIAAAPPSPVTVGSSLSSVLGPPVPEIKVKEEVEPMDIMRP 796
 P + +T P +PP TT PPP+ T ++ + PP + P P
 Sbjct: 1589 PTITTTTTPPTTTPSPPTT-TTTPPTTTPSPPTTTPITPPTSTTTLPPTTTPSP--PP 1645

Query: 797 VSAVPPLATNTVSPSLALLANNLSMPTSDLPFGASP 832
 + P T T SP + T+ PP +P
 Sbjct: 1646 TTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTP 1681

Score = 183 (27.5 bits), Expect = 3.4e-09, P = 3.4e-09
 Identities = 91/390 (23%), Positives = 139/390 (35%)

Query: 326 PSAAISIQRPAQSRDVTTR-ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPIL 384
 PS + P + T T PS P T T I +T TP+ T +P +
 Sbjct: 1399 PSPPTTTPSPPTTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPSPPI 1458

Query: 385 ATNTIPSAATTAGSVSHTQAPTSTIVTMTVPSHSHATAVTTSNIP--VAKVVQIQITHS 442
 +T T P TT S T P+ T + P+ ++ TT+ P + P T T
 Sbjct: 1459 STTTTPPTTTPSPPT-TTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPITPASTTL 1517

Query: 443 PRIQPDYPAERSSSLIPISGHRASP---NPVAMETRSDNR--SVPVQFQYFLPTYPPSAY 497
 P P ++ P SP P+ T + P + P T PP+
 Sbjct: 1518 PPTTTPSPPTTTPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPTTTPPTTTPPTT 1577

Query: 498 PLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNMQ-LMTVDASHAR 556
 P T TP ++T P + +P T T +T P +T S
 Sbjct: 1578 PSPPTTTPSPPTITTTTPPTTTPSP--TTTTPPTTTPSPPTTTPITPPTSTTT 1634

Query: 557 HIQGIQAPISTQGIQAPIGTPGIQAPPLGTQGIHSATPINTQGLQAPMGTQQPQPEG 616
 P+P T P P TP P P T T P P
 Sbjct: 1635 LPPTTTPSPPTTTPPTTTPS--P-PTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPSP 1691

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPA 676
 T+ + T ++PI+ + P+TT + +T +P SP + + P
 Sbjct: 1692 PTTTMTTPSTTTPSPPTT-TTTPSTTTPSPPTTMTTPSPPTTTPSPPTTMTTLPP 1749

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPP 715
 T + P + + P +++ T S + PT P
 Sbjct: 1750 TTTSSPLT----TTLPPSITPPTTFSPSTTTPTTPCVP 1784

Score = 176 (26.4 bits), Expect = 1.8e-07, P = 1.8e-07
 Identities = 101/402 (25%), Positives = 142/402 (35%)

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATNTIPSAATTAGSVSHTQAP 404
 IT PS P TP T +T +P T P T P TT + T P
 Sbjct: 1396 ITTPSPPTT-TPSPPTTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTP 1454

Query: 405 TSTIVTMTVPSHSHATAVTT-SNIPVAKVVQIQITHSPRIQPDYPAERSSSLIPISGHR 463
 + I T T P ++ + TT+ + P P T T+P P PI+
 Sbjct: 1455 SPPISTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1511

Query: 464 ASPNPVAMETRSDNRSPVPVQFQYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQA 523
 AS + T PS P TP+ P + T TPIT ST P + +
 Sbjct: 1512 ASTTTLPTTT----PSPPTTT---TTPPTTTP-SPPTTTPITPPTSTTTLPPTTTPS 1563

Query: 524 PNSAITAQ---TGVGVAHVHLNMQMTVDASHARHIQGIQAPISTQGIQAPIGTP 579
 P T T +T +P +T P+P +T P P TP
 Sbjct: 1564 PPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1618

Query: 580 G-----IQAPPLGTQGIHSAT---PINTQGLQAPMGTQQPQPEGKTSAVVLADGATIV 630
 I P T + T P T P P T P S +
 Sbjct: 1619 SPPTTTPITP-PTSTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTP 1677

Query: 631 ANPISNPFSAAPAA-TTVVQTHSQSASTNAP-AQGSSPRPSILRKKPATDGAKPKSEIHV 688
 S+P + P+ TT + T S + + ++P ++P + P T P
 Sbjct: 1678 TTTSSPITTTTSPPTTTMTTTPSPTTTPSSPITTTTTTSSSTTTPSPPTTMTTTPSP---T 1734

Query: 689 SMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPSPQFAVALSTIPG 746
 + +P T +M T+ P P PPT + + P+ P V L G
 Sbjct: 1735 TTPSPPTTMTTLPPTTTSSPLTTTLPSPITPPTFSFP--STTTPTTCVPLCNWTG 1790

Score = 168 (25.2 bits), Expect = 9.3e-08, P = 9.3e-08
 Identities = 89/387 (22%), Positives = 133/387 (34%)

Query: 448 DYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFYFLPTYPPSAYPLAHTYTP 507
 DY + P+ +P+P T + + P P PT PS P T P
 Sbjct: 1381 DYKIRVNCCWPMCKCITTPSP---PTTTPSP--PTTTTLPPTTTPSP-PTTTTTPPP 1434

Query: 508 TSSVS---TIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQPA 564
 T++ S T P+ P+ I+ T +T P T + P+
 Sbjct: 1435 TTTSPPTTTTTLPTTTPSPPTTTPPTTTT---PSPPTTTPSPPTT-----TPS 1485

Query: 565 PISTQGIQAPIGTPTGI-QAPPLGTQGIHSATPINTQGLQAPAPMGTOQPQ---PEGKTS 620
 P +T P P TP P+ + P T P T P P T+
 Sbjct: 1486 PPTTTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTT 1545

Query: 621 VVLADGATIVANPISNPFSAAPAAATTVVQTHSQSA-STNAPAQGS---SPRPSILRKKP 675
 + +T P + P TT T + S +T P+ + +P P+ P
 Sbjct: 1546 PITPPTSTTTLPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSP 1605

Query: 676 ATDGAKPKSEIHVS--MATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASP 733
 T P S TP+T T + P+ P T PPPT +
 Sbjct: 1606 TTTTTPPTTTPSPPTTTPITPPTSTTTLPTTTPS-PPPTTTPPTTTPSPPTTTP 1664

Query: 734 PSQPAVALSTIPGAVPITPPTITIAAAPPSVTVGSSLSSVLGP---PVPEIKVKEEVE 789
 PS P +T P + PITT + P ++T ++ P P
 Sbjct: 1665 BSPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPSP 1724

Query: 790 PMDIRPVSAVPLATNTVSPSLALLANLNSMPTSDLPFGASP 832
 P + P P T +L + + T+ LPP +P
 Sbjct: 1725 PTTMTTPSPPTTTPSPPTTTPPTTTPPTTSSPLTTTLPSPITP 1767

Score = 154 (23.1 bits), Expect = 2.7e-06, P = 2.7e-06
 Identities = 70/277 (25%), Positives = 92/277 (33%)

Query: 565 PISTQGIQAPIGTPTGIQAPPLGTQGIHSATPINTQGLQAPAPMGTOQPQPEGKTS 624
 PIST P P TP P P T + TP P T P P T +
 Sbjct: 1457 PISTT-TTPPTTTPS--P-PTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTP 1510

Query: 625 DGATIVANPISNPFSAAPAAATTVVQTHSQSASTNAP----AQGSSPRPSILRKKPATDGA 680
 +T P + P TT T + S T P ++ P+ P T
 Sbjct: 1511 PASTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPPTT 1570

Query: 681 KPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQ--PPPTIPTMIAAASPSPQFA 738
 P S T T S T++ T PPT PPPT T + P P
 Sbjct: 1571 TTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1629

Query: 739 VALSTIPGAVPITPPTITIAAAPPSVTVGSSLSSVLGPVPVEIKVKEEVEPMDIRPV 798
 + +T+P +PP TT PPP+ T ++ PP+ +
 Sbjct: 1630 TSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSP 1688

Query: 799 AVPLATNTV-----SPSLALLANL--SMPTSDLPFGASPRKKP 836
 PP T T +PS + S T PP P
 Sbjct: 1689 PSPPTTMTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1733

Score = 148 (22.2 bits), Expect = 1.1e-05, P = 1.1e-05
 Identities = 62/254 (24%), Positives = 89/254 (35%)

Query: 583 PAPLGTQGIHSATPINTQGLQAPAPMGTOQPQPEGKTSAV-----VLADGATIVANPISNP 637
 P+P T S P T L P T P P T+ + T P+
 Sbjct: 1399 PSPPTTTP--SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1452

Query: 638 FSAAPAAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKPKSEIHVS--MATPVT 695
 + P +TT T + + + P SP P+ P T P S M TP+T
 Sbjct: 1453 TSPPTTSTTT--TPPTTTPSPPTTTPSP-PTTTPSPPTTTPPTTTPPTTTPSP 1509

Query: 696 VSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPSPQFAVALSTIPGAVPITPPT 755
 T + P+ T PP T P+ + P P + +T+P +PP T
 Sbjct: 1510 PPASTTTLPTTTPSPPTTTPPTTTPPTTTPS--PPTTTPITPPTSTTTLPTTTPSPPT 1567

Query: 756 TIAAAPPSVTVGSSLSSVLGPVPVEIKVKEEVEPMDIRPVSAVPLATNTVSPSLALL 815
 T PPP+ T ++ PP + PP T P+ +
 Sbjct: 1568 T-TTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPPTT 1626

Query: 816 ANNLSMPTSDLPFGASPRKKP 836
 S T+ LPP +P P
 Sbjct: 1627 TPPTS--TTTLPPTTTPSPPP 1645

Score = 131 (19.7 bits), Expect = 1.2e-03, P = 1.2e-03
 Identities = 112/492 (22%), Positives = 174/492 (35%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3977 VTPTPTGTGTPTTTTPTTTTTVTPTPTGTGTPTTTTPTTTTTVTPTPTGTGTPT 4036

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTNNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 4037 TTPITTTTTVTPTPTGTGTPTTTTPTTTTTVTPTPTGTGT-PTTTPTTTTTVTPT 4095

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 4096 PTPGTGTPTTTTPTTTTTVTPTPTGTGTPTTTTPTTTTTVTPTPTGTGTPTTTTPI 4155

Query: 269 IHQPIQSRPPVTTNSAIPPA--VVATVSATRAQSPVITTTA--AHATDSALSRPTLSIQH 324
 + P T P + T +T +P T T H + + + T S
 Sbjct: 4156 TTTTPTPTPTGTGTPTTTTPTTTTTVTPTPTGTGTGPPHTSTAPIAELTTSNPP 4215

Query: 325 PPSAAISIQRPQS--RDVTTRI-TLPSHPALGTPKQQLHTMAQKTI FSTGTPVAAATVA 381
 P S+ R S + TT + TLP PA+ + T T + T T+
 Sbjct: 4216 PESTPTQSRSTSSPLTESTTLLTLP--PAI----EMTSTAPPSTPTAPTTSGGHTLS 4269

Query: 382 PILATNTIPSAT-TAGSVS-HTQAPTSTIVTMTVPSHSHATAVTTSNIPVAKVVPQIT 439
 P +T T P T T G+ + + APT + V T S A T + P++ P I
 Sbjct: 4270 PPPSTTTSPPGTPTRGTTTGSSSAPTSTVQTTTT--AWTPTPTPLS--TPSIIR 4321

Query: 440 HTSPRIQPDYPAERSSSLIPISGHRASPNP-VAMETRSDN----RPSVPVQFYFLPTYF- 493
 T ++P YP+ ++ +P V T D S+ +++ + P
 Sbjct: 4322 TTG--LRP-YPSSVLICCVLNDTYYPAGEEYNGTYGDTCTYFVNCSLSCLEFYNNWSCPS 4378

Query: 494 -PSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDA 552
 PS P + + TP S S+ P P T L + T
 Sbjct: 4379 TPSPTTPSKS-TPTPSKPSSTPSKPTPGTKPPECDFDPQRQENETWWLDCDFMATCKY 4437

Query: 553 SHARHIQGIQ----PAPISTQGIQPAIGTP 579
 ++ I ++ P P + G+Q P + P
 Sbjct: 4438 NNTVEIVKVECEPPMPTCSNGLQPVREDP 4468

Score = 117 (17.6 bits), Expect = 1.8e-02, P = 1.8e-02
 Identities = 41/156 (26%), Positives = 55/156 (35%)

Query: 710 TIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPSVTVGG 769
 T + P T PPPT T + + PS P +T P +PPITT P P+ T
 Sbjct: 1398 TPSPTTTSPPTTTTTLPTTTTSPPTTTTTPPTTTTSPPIIT-TTTLPTTTTSP 1456

Query: 770 SLSSVLGPPVPEIKVKEEVEPMDIRPVSAVPLATNTVSPSLALLANNLSMPTSDLPFG 829
 +S+ PP P P + P T T SP T+ PP
 Sbjct: 1457 PISTTTTTP-----PTTTPSPPTTTPSPPTTTPSPPTTTTTP-PTTTTSPPM 1504

Query: 830 ASPRKKPRKQHVISTEEDMMETNSTDDEKSTAKS 865
 +P P + T T+T +T S
 Sbjct: 1505 TTPITPPASTTTLPTTTTSPPTTTTTPPTTTTTPS 1540

Score = 61 (9.2 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 23/93 (24%), Positives = 41/93 (44%)

Query: 397 SVSHTQAPTSTIVTMTVPSHSHATAVTTSNIPVAKVV----PQOITHTSPIQPDYPAE 452
 S++ + +T T+T+P+ + T TT+ P + V P+ S I D+P+
 Sbjct: 1257 SITTRPSTLTFTTTITLPTTPTSFTTTTTTTTPTSSVTLSTTPKLCLLWSDWINEHPSS 1316

Query: 453 RSS---LIPISGHRASPNPVAMETRSDNRPSVPVQ 484
 S P G +P + E RS P + ++
 Sbjct: 1317 GSDDGDREPFDFGVCAPEDI--ECRSVKDPHLSLE 1349

Score = 50 (7.5 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 16/41 (39%), Positives = 19/41 (46%)

Query: 334 RPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTP 374
 RP+ TT ITLP+ P T T T+ ST TP
 Sbjct: 1261 RSTLTFTTT-ITLPTTPTSFTTTTTTTTPTSSVTLST-TP 1299

Score = 46 (6.9 bits), Expect = 5.4e-08, P = 5.4e-08
 Identities = 24/106 (22%), Positives = 37/106 (34%)

Query: 324 HPPSAAISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTPVAAATVAPI 383
 +PP A++ + +S T + P G Q A G I

Sbjct: 1196 YPPGASVPTEETCKSCVCTNSSQVVCPRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNI 1255

Query: 384 LATNTIPSA-TTAGSVSHTQAPTSTIVTMTVPSHSHATAVTTNSI 428
 + T PS TT +++ PTS T T + +S TT +
 Sbjct: 1256 CSITTRPSTLTFTTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKL 1301

Score = 44 (6.6 bits), Expect = 8.7e-08, P = 8.7e-08
 Identities = 14/34 (41%), Positives = 17/34 (50%)

Query: 478 RPSVPVQFQYF-LPTYPPSAYPLAHTYTPITSSV 511
 RPS F LPT P S + T TP +S+V
 Sbjct: 1261 RPSTLTFTTTITLPTTPTS-FTTTTTTTTPTSSTV 1294

Pedant information for DKFZphtes3_2all, frame 2

Report for DKFZphtes3_2all.2

[LENGTH] 1048
 [MW] 110324.04
 [pI] 9.83
 [HOMOL] PIR:I47141 gastric mucin (clone PGM-2A) - pig (fragment) 8e-15
 [FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-09
 [FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-09
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-09
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR420w] 4e-09
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR420w] 4e-09
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR151c] 4e-06
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGR014w] 1e-05
 [FUNCAT] 11.01 stress response [S. cerevisiae, YHL028w] 1e-04
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YHL028w] 1e-04
 [EC] 3.2.1.3 Glucan 1,4-alpha-glucosidase 3e-08
 [PIRKW] glycosidase 3e-08
 [PIRKW] transmembrane protein 3e-08
 [PIRKW] polysaccharide degradation 3e-08
 [PIRKW] glycoprotein 9e-08
 [PIRKW] calcium binding 9e-08
 [PIRKW] hydrolase 3e-08
 [PIRKW] cytoskeleton 7e-08
 [SUPFAM] equine herpesvirus glycoprotein X 2e-07
 [SUPFAM] yeast glucan 1,4-alpha-glucosidase homolog 3e-08
 [SUPFAM] polymorphic epithelial mucin 7e-08
 [SUPFAM] glucan 1,4-alpha-glucosidase homology 3e-08
 [SUPFAM] equine herpesvirus 1 glycoprotein homology 2e-07
 [PROSITE] MYRISTYL 9
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 10
 [PROSITE] PKC_PHOSPHO_SITE 12
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] Irregular
 [KW] LOW_COMPLEXITY 20.04 %

SEQ MGPPRHPQAGEIEAGGAGGRRRLQVEMSSQQFPRLGAPSTGLSQAPSOIANSAGSAGLINP
 SEGXXXXXXXXXXXXX.....
 PRD ccc

SEQ AATVNDESGROSEVSAREHMSSSSSLSREEKQEPVVVRPYPQVQMLSTHHAVASATPVA
 SEGXXXXX.....XXXXXXXXXXXXX.....
 PRD ccc

SEQ VTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPGQVTVTMESSI
 SEG XXXXXXXXXXXXXXXX.....XXXXXXXXXXXXX.....
 PRD ccc

SEQ PQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPPLHIGASHLPRGAAAAAVM
 SEGXXXXX.....
 PRD ccc

SEQ SSSKVTTLRLPTSQPLNAATAQPAVQHIHQIQRPPVTTNSAIPPAVVATVSATRAQS
 SEG
 PRD ccc

SEQ PVITTTAAHATDSALSRLPTLSIQHPPSAAISIQRPAQSRDVTTRITLPSHPALGTPKQQL
 SEG
 PRD ccc


```

SEQ HTMAQKTI FSTGT PVAAATVAPILATNTIPSATTAGSVSHQTAPTSTIVTMTVPSSHSHA
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TAVTTSNI PVAKVVPQOITHTSPIRQPDYPAERSSLIPI SGHRASPNPVAMETRSDNRPS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ VVPQFQYFLPTYPSPAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HLNPMQLMTVDASHARHIQGIQFAPISTQGIQFAPIGTPIGQFAPLGTQGIHSATPINTQ
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GLQPAPMGTQQPQPEGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AQGSSPRPSILRKKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPPTAQQP
SEG .....XXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPSPVTVGGSLSSVLGPPVP
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EIKVKEEVEPMDIMRPVSAVPPPLATNTVSPSLALLANLSMPTSDLPFGASPRKKPRKQQ
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HVISTEEGDMMETNSTODEKSTAKSLLVKAERKSPKEYIDEEGVRYVPVRPRPPITLL
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RHYRNPWKAAYHHFQRYSDVRVKEKKAMLQEIANKQGVSCRAQGWKVLCAAQLLQLTN
SEG .....
PRD ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LEHDVYERLTNLQEGII PKKKAATDDDLHRINELIQGNMQRCKLVMDQISEARDSMLKVL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DHKDRVLKLLNKNGT VKKVS KLKRKEV
SEG .....XXXXXXXXXXXX
PRD hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_2a11.2

PS00001	818->822	ASN_GLYCOSYLATION	PDOC00001
PS00001	854->858	ASN_GLYCOSYLATION	PDOC00001
PS00001	1033->1037	ASN_GLYCOSYLATION	PDOC00001
PS00004	872->876	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1037->1041	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005
PS00005	242->245	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	665->668	PKC_PHOSPHO_SITE	PDOC00005
PS00005	831->834	PKC_PHOSPHO_SITE	PDOC00005
PS00005	862->865	PKC_PHOSPHO_SITE	PDOC00005
PS00005	940->943	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1035->1038	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	88->92	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	473->477	CK2_PHOSPHO_SITE	PDOC00006
PS00006	844->848	CK2_PHOSPHO_SITE	PDOC00006
PS00006	855->859	CK2_PHOSPHO_SITE	PDOC00006
PS00006	959->963	CK2_PHOSPHO_SITE	PDOC00006
PS00006	984->988	CK2_PHOSPHO_SITE	PDOC00006
PS00008	15->21	MYRISTYL	PDOC00008

PS00008	16->22	MYRISTYL	PDOC00008
PS00008	36->42	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	535->541	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	768->774	MYRISTYL	PDOC00008
PS00009	19->23	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_2a11.2)

DKFZphtes3_2a17

group: metabolism

DKFZphtes3_2a17 encodes a novel 574 amino acid protein without similarity to known proteins.

The novel protein contains a thiol protease cys pattern. Eukaryotic thiol proteases (EC 3.4.22.-) are a family of proteolytic enzymes containing an active site cysteine. Cathepsins belong to this protease family.

The new protein can find application in modulation of proteolytic processes and as a new enzyme for proteomic analysis and biotechnologic production processes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 2312 bp

Poly A stretch at pos. 2300, polyadenylation signal at pos. 2273

```

1  GTTTCACCT  GATCATTAGA  AACTAATGAA  ACACCTTTTA  AGTCTTATGA
51  ATTCAAGGTTA  CACTGTTTTTC  CAGATGCCTT  GGCAGCTGGT  ACAGGGCCTC
101  TGAAAAATGG  AACCAAAATTC  TCTGAGGACT  AAAGTCCCAG  CTTTCTTATC
151  TGATTTGGGG  AAGGCCACAT  TGAGGGGAAT  CAGAAAGTGT  CCCCAGTGTG
201  GCACATACAA  TGGAACCCGG  GGAAGTGAAT  GTAAGAACA  GACATGTGGA
251  ACCATATTCC  GCTACGGTGC  ACGCAAGCAG  CCTAGTGTG  AAGCTGTCAA
301  AATCATTACA  GGCTCTGATC  TTCAGGTCTA  CTCAGTGGCG  CAAAGAGACC
351  GGGGCCCTGA  TTACCGATGC  TTTGTGGAGC  TCGGGGTTTC  AGAGACAACA
401  ATCCAGACAG  TGGATGGGAC  GATCATCACT  CAGCTGAGCT  CTGGACGGTG
451  TTAATGTCCT  TCATGCCTGA  AAGCTGCCAC  TCAAGGCGTT  GTGGAAAACC
501  AGTGCCAGCA  CATCAAGCTG  GCGGTGAAGT  GCCAGGCAGA  GGCCACCCCT
551  CTGACCTGTA  AGAGCTCGGT  CCTGAATGCA  ATGCAGGCCT  CCCCAGAAAC
601  CAAACAGACC  ATCTGGCAGT  TGGCCACGGA  ACCCAGAGGT  CCTCTGGTGC
651  AGAGAATTAC  TAAAAACATC  TTGGTGGTGA  AATGCAAGGC  AAGCCAGAAG
701  CACAGTTTGG  GGTATTTGCA  TACATCTTTT  GTGCAGAAAG  TCAGTGGCAA
751  AAGCTTGCTC  GAGCGCCGCT  TCTTCTGCTC  CTGTCAGACT  CTGAAATCGC
801  ACAAGTCAAA  TGCCCTCCAAG  GATGAGACAG  CCCAGAGATG  CATTCATTTT
851  TTTGCTTGCA  TCTGTGCCTT  TGCCAGTGAT  GAGACACTGG  CTCAGGAATT
901  CTCAGACTTC  CTAAATTTTG  ATTCCAGCGG  TCTTAAAGAG  ATTATTGTAC
951  CCCAGTTAGG  TTGCCATTCA  GAATCAACAG  TATCTGCTTG  TGAGTCTACT
1001  GCCTCTAAGT  CAAAGAAGAG  GAGAAGGAT  GAAGTATCTG  GTGCACAGAT
1051  GAACAGTTCA  CTACTGCCTC  AAGATGCAGT  GAGCAGTAAT  CTAAGGAAAA
1101  GTGGCCTGAA  AAAGCCTGTG  GTTGCTTCCT  CGTTAAAAAG  GCAGGCCCTG
1151  GGTGAGCTGT  TAGATGAGGC  ACAAGTGACT  TTATCCTTCC  AAGACTGGCT
1201  GGCCAGTGTC  ACAGAACGCA  TCCATCAAAC  CATGCATAT  CAGTTTGATG
1251  GCAAACAGCA  ACCATTGGTG  TTCCACATTC  CTCAGTCATT  TTTTGATGCC
1301  CTGCAACAAA  GAATATCTAT  AGGAAGTGCA  AAAAAACGGC  TCCCCAACTC
1351  CACCACAGCT  TTTGTTGCGA  AAGATGCCTT  GCCACTGGGA  ACCTTTTCCA
1401  AGTATACTTG  GCATATCACT  AATATCCTGC  AAGTTAAACA  AATCTTAGAT
1451  ACCCCAGAGA  TGCCCTTGGA  AATCACCCTG  AGCTTTATCC  AGAACCGAGA
1501  TGGGACTTAT  GAGCTATTTA  AATGCCCTAA  AGTGGAAAGT  GAAAGCATAG
1551  CAGAAACCTA  CGGTGCTATA  GAAAAACAAC  CAGTGTGCG  ACCCTTGGA
1601  CTAAAACTT  TTCTCAAAGT  TGGCAACACT  TCCCAGATC  AAAAGGAGCC
1651  AACACCTTTC  ATCATCGAGT  GGATCCCAGA  TATCCTTCCC  CAATCTAAGA
1701  TTGGCGAGCT  GCGGATCAAG  TTTGAGTATG  GCCACCACCG  GAATGGGCAT
1751  GTGGCGGAGT  ACCAAGACCA  GCGGCCCCCT  TTGGACCAGC  CCTTGGAAC
1801  GGCCCTCTG  ACCACTATTA  CTTTCCCTTA  AAGCAAAACA  AGATAATAAT
1851  CTTTGTGCTG  TTAATTGCA  CATCCCCACC  CCTTGACAAC  TTTAAATGCT
1901  AGTTAGGCAC  TTAGATGGCC  CTGTTCTTTG  GTAAACTGCT  CTTAGCTAAG
1951  ATGCAAAATC  TCAGTGCTTT  CAAGTGGATT  CTGTTGAAGA  AAATCTCTTG
2001  TAAATAGCCT  TTTTGATGCT  GCTGTGTACA  GTCTTCATTA  TGCATTGGGC
2051  AGTATTTCTG  GCTAGAGTTT  TAAAAGGAAC  AGAAAGAAAA  CCAGCTTATT
2101  TTCTTCTTGA  CGGACTCATC  TTTAGCGTTT  ATTTCAACCT  TTTGCTAATT
2151  CTCTGAGAAA  TCTGCAGCAC  TCAGCCATAC  ACCAACAGTG  TTGGAAAGTT
2201  AACACCCCTG  TTAGGCAGCA  ATGTTAAAGA  CCATCTTGCC  AGAGTTCCAG
2251  CCACGCTCTT  TATTCTGTTC  TCAATAAAG  CAGTGTCACT  AGTTTTTCTT
2301  AAAAAAAAAA  AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 1828 bp; peptide length: 574
 Category: putative protein

```

1 MEPNSLR TKV PAF LSDLGKA TLRGIRK CPR CGTYNGTRGL SCKNKTCGTI
51 FRYGARKQPS VEAVKIITGS DLQVYSVRQR DRGPDYRCFV ELGVSETTIQ
101 TVDGTIITQL SSGRCYVPSC LKAATQGVVE NQCQHIK LAV NCQAEATPLT
151 LKSSVLNAMQ ASPETKQTIW QLATEPTGPL VQRITKNILV VKCKASQKHS
201 LGYLHTSFVQ KVS GKS LPER RFFCSCQTLK SHKSNASKDE TAQRCIHFFA
251 CICAFA SDET LAQEFSDFLN FDSSGLKEII VPQLGCHSES TVSACESTAS
301 KSKKRRKDEV SGAQMNSSLL PQDAVSSNLR KSGLKKPVVA SSLKRQACGQ
351 LLDEAQV TLS FQDWLASVTE RIHQTMHYQF DGKPEPLVFH IPQSFFDALQ
401 QRISIGS AKK RLPNSTTAFV RKDALPLGTF SKYTWHITNI LQVKQILDTP
451 EMPLEITRSF IQNRDGTIEL FKCPKVEVES IAETYGRIEK QPVLRLPLELK
501 TFLKVGNTSP DQKEPTFFII EWIPDILPQS KIGELRIKFE YGHHRRNGHVA
551 EYDQQRPPLD QPLELAPLTT ITFP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2a17, frame 2

Report for DKFZphtes3_2a17.2

```

[LENGTH]      574
[MW]           64076.89
[pI]           9.15
[PROSITE]      MYRISTYL           5
[PROSITE]      CK2_PHOSPHO_SITE    9
[PROSITE]      PKC_PHOSPHO_SITE   14
[PROSITE]      ASN_GLYCOSYLATION   5
[PROSITE]      THIOL_PROTEASE_CYS  1
[KW]           Alpha_Beta

```

```

SEQ  MEPNSLR TKVPAFLSDLGKATLRGIRK CPRCGTYNGTRGLSCKNKTCGTIFRYGARKQPS
PRD  cccccccccchhhhhccccchhhhhcccccccccccccccccccccccccccccccccccc

```

```

SEQ  VEAVKIITGS DLQVYSVRQRDRGPDYRCFV ELGVSETTIQTVDGTIITQLSSGRCYVPSC
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

```

SEQ  LKAATQGVVENQCQHIK LAVNCQAEATPLTLKSSVLNAMQASPETKQTIWQLATEPTGPL
PRD  hhhhhhhhhccccchhhheehhhhhhhccccchhhhhhhhhccccchhhhhhhhhccccch

```

```

SEQ  VQRITKNILVVKCKASQKHS LGYLHTSFVQKVS GKS LPERFFCSCQTLKSHKSNASKDE
PRD  hhhhhhhheeeeecccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  TAQRCIHFFACICAFASDETLAQEFSDFLNFDSSGLKEIIVPQLGCHSESTVSACESTAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  KSKKRRKDEVSGAQMNSSLLPQDAVSSNLRKSGLKKPVVASSLKRQACGQLLDEAQV TLS
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

```

SEQ  FQDWLASVTERI HQTMHYQFDGKPEPLVFHI PQSFFDALQQRISIGS AKKRLPNSTTAFV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  RKDALPLGTF SKYTWHITNILQVKQILDTPEMPLEITRSFIQNRDGTIELFKCPKVEVES
PRD  eccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

```

SEQ  IAETYGRIEKQPVLRLPLELKTFLKVGNTSPDQKEPTFFII EWIPDILPQSKIGELRIKFE
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```


SEQ YGHRNRNGHVAEYQDQRPPLDQPLELAPLTTITFP
 PRD ecccccccccccccccccccccccccccccccccccc

Prosites for DKFZphtes3_2a17.2

PS00001	35->39	ASN_GLYCOSYLATION	PDOC00001
PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	235->239	ASN_GLYCOSYLATION	PDOC00001
PS00001	316->320	ASN_GLYCOSYLATION	PDOC00001
PS00001	414->418	ASN_GLYCOSYLATION	PDOC00001
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	231->234	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	369->372	PKC_PHOSPHO_SITE	PDOC00005
PS00005	407->410	PKC_PHOSPHO_SITE	PDOC00005
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	360->364	CK2_PHOSPHO_SITE	PDOC00006
PS00006	367->371	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	508->512	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	312->318	MYRISTYL	PDOC00008
PS00139	109->121	THIOL_PROTEASE_CYS	PDOC00126

(No Pfam data available for DKFZphtes3_2a17.2)

DKFZphtes3_2d15

group: testes derived

DKFZphtes3_2d15 encodes a novel 274 amino acid protein with similarity to *C.elegans* cosmid F25H2.1.

The novel protein contains a Pfam predicted C2-domain.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *C.elegans* F25H2.1

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3615 bp

Poly A stretch at pos. 3603, polyadenylation signal at pos. 3578

```
1 GCGGCGGCCCT CGAGGTGACA ACTGTCTCCG TCGCAGGCTC CGGCGGGGGC
51 GCAGGAGGTC GCCCGGCGCG TCACTGTCCG GTCGGCGAGC CACGGGGGGC
101 GCCGCAGCAC CATGGCGACC ACCGTACAGCA CTCAGCGCGG GCCCGTGTAC
151 ATCGGTGAGC TCCCGCAGGA CTCCTCCCGC ATCAGGCCCA CACAGCAGCA
201 GCGGCAGGTC CAGCTGGACG CCCAGGCGGC CCAGCAGCTG CAGTACGGAG
251 GCGCAGTGGG CACCGTGGGC CGACTGAACA TCACGGTGGT ACAGGCAAGG
301 TTGGCCAAGA ATTACGGCAT GACCCGCATG GACCCCTACT GCCGACTGCG
351 CCTGGGCTAC GCGGTGTACG AGACGCCAC GGCACACAAT GCGCCAAGA
401 ATCCCCGCTG GAATAAGGTC ATCCACTGCA CGGTGCCCCC AGGCGTGGAC
451 TCTTCTATC TCGAGATCTT CGATGAGAGA GCCTTCTCCA TGGACGACCG
501 CATTGCCCTG ACCCACATCA CCATCCCGGA GTCCCTGAGG CAGGGCAAGG
551 TGGAGGACAA GTGGTACAGC CTGAGCGGGA GGCAGGGGGA CGACAAGGAG
601 GGCATGATCA ACCTCGTCAT GTCCTACGCG CTGCTTCCAG CTGCCATGGT
651 GATGCCACCC CAGCCCGTGG TCCTGATGCC AACAGTGTAC CAGCAGGGCG
701 TTGGCTATGT GCCCATCACA GGGATGCCCG CTGTCTGTAG CCCCGGCATG
751 GTGCCCCGTG CCCTGCCCCC GGGCGCCGTG AACGCCACG CCCCGTGTAG
801 CGAGGAGGAC CTGAAAGCCA TCCAGGACAT GTTCCCAAC ATGGACCAGG
851 AGGTGATCCG CTCCTGTGTC GAAGCCACG GAGGGAACAA GGATGCCGCC
901 ATCAACTCCC TGCTGCAGAT GGGGGAGGAG CCATAGAGCC TCTGCCTCGA
951 TGCCGTTTTG CCCCCTCTCT TTGGACACGC CGACCCGGCG CTCCCAAGG
1001 AATGCTGTCC CAACAAGATT CCCGTGAAAG AGCACCCTG TCGCCCCCTC
1051 CCGTGGACTT CTGTGCCGCC CCGTCCACAC CTGTTCTTGG GTGCATGTGG
1101 GTTTTCGGTT CCTGGCGGTC CAGGACGGGG CGGGGGCTCC CCTCCCATCT
1151 CGTGCTGGGA GGTCTCAGCG CGCTCTCCTG TCCTTGGGAC GTGCGTCTCT
1201 CCTTCTCATG CCGTTCTGGA AAATGCTCTT GCTGTAGAGA GCAGTGCTT
1251 CTGCCAGGTT GTTGGAGGTG GTGGAGCGCC TTCCGATTCC ATTCATGGCA
1301 TTTTGTGATG TGATGTAATT GGAATAGAGC TGTGATTATA AGGCACACAC
1351 AATCCCTCAC ACTGTGGGTT TTTTGTAGAA CTCCAGAC GAAAACTCAC
1401 GCCCTTGCCC TAACGCGCTT TGCTGTGAGC CTGGCCCTG CCCAGGGCTT
1451 GGGTCTGTGT AGCTGAGCAG CTCTCTGTGG ATGGTGTGGG GCGGGCTCT
1501 GGCCTGGGTC ACCTGGCCAC TGTCCAGCCA GCCTTGTGAC AGACTCCGGC
1551 CTGAAGGCAG AATGAACCCA CACCTGGAGT GAGGAAGGGG GCCTGGCACG
1601 GTTGGCCAGG CTCTGCCTGA TTGCCAGCCA GCGGGCATCT GAAGCCGGGT
1651 CCTTCGCCCG CCGGAGGCTG CCGTCCGTCT CTCCTGCTGC GCTCGTGCCA
1701 GCTCCGTGGG TGTCCTCCA GGGAGCTTCT CTCTCAACA GGCCTTGCGA
1751 GGCTGGGGTG AGAGGTGATA GAGGCAGCAC TGTGCATGAT TCCGAGAGGG
1801 TGTGGTGGCA CTGCCAGCCG ACTGCTGACA GCTTGGGAGC TGCTGTGCCC
1851 AGGACGTGGG TTCAGCGTGG GCGAGGAAAG CCTGGCGAGC GTGGCCCTGT
1901 AAAAGCTTTC TGAGGCGGGA GCGCTCACT TACCTGTGAC TGCTGGGCG
1951 CTGCGGTGTA CATCTTGGCC TACAGGACAG ATTTTAGGTG ACACCTGGTT
2001 ATGACAGTCA GAAATTGAG AAGCTTCTCA CAAGTGATGC ACTTTAAATA
2051 ATCTGCATGC CATTGAGACA CCTGCATGTC TGGTGTGTGT GGTTCAGTG
2101 TCTTGCCGCC GGCCTTCGGA TGTAAACCCA CTGATAACCG ACAGAAAGAG
2151 AATGCCACCA AGTGGGTCTT CTGTGGAAGA TGCAGAAGGA GGAAGTTAGT
2201 GCTTACATTT TAGTCTTTT CTCCCTCAAA AAAATAGGTT AAGTTTCAGT
2251 GCCAGCTAGA AAATACTGCT TTCTGCCATC GATTGGGGGT GGTTTTGTG
2301 AAATATACTG TTGATAAATA TTTATTTTGT TAAACTTGAA GTGTGTGGTG
2351 GCCGTGGGGG AGGGACATGC TGGCAGCAGG CGCCTTCTTC AGCTGTGGGT
2401 CCTAAAGGCC TTTGATCCTT TGAAGAAGAA AGACATGGTA TTTGTTGAGC
2451 AGAGCCCGAC CACTCAGACG GAGGGGCCCC TGGGATTCCC TGCTCTCAGT
2501 GCGCTGGGCT TACGCCTGTG TAGATTCTTT CTCCATTGGG AATGAAGGTG
2551 TCAGGCGGGA CTGGAACGTT CTAGATGGTA TGTCCCGTGA TATTAACAA
2601 TCTAACCCAG GACAGACCAC AAGCCACACT CAGAGGCCCT ACTGTGCTTG
```



```

2651 GGGCTTCGGT GTCCAGGCGC CCAGGTGTGG CCACCAGCAC CGGTTTCTGC
2701 CTTCCGCTTG CTGGGGTGCA GTGAGACTGC CACACGCGTG CACATGTGGC
2751 TCTGTGGGTG TCTCCTAGAG AGGACGTGGC CCCTGCTGCC AGCCCTTGAG
2801 CAGCCCCGTGT GGGGGCCCGA GGGACCCACA CAGTGGGGGC CAGCCTCGCT
2851 GGAGGGAGAG CAACCCCTTG CCGATGACCA CGCTTGCCGC CATCTCTTAG
2901 TTTTCTTTTT CACAAGCGCT TTATTTTTTT AATAGACAAA TCACATTTTG
2951 CAAGGCCTTT AATTAAATAA GATTCTTCTT TCCTTCATTT TATGCTTTAT
3001 TTCTGTGTTG AAGGCTTACT GTAGAAGTGG CTTACTGTAG AAGCAGCTTG
3051 CTGAGCCCTT CCGAGCGGTC CCCAGAATTA GCTGGTTCAC AACCCCCACC
3101 CTCCTCCGCT CCCGCTGTG TCAGGTGTGG ATGAGGTCGT CACACTCAGA
3151 AGGACAGGCT TGTCTGCCAG CTCACAAGGG GAGGCTGCAG TGGGTTTGGG
3201 AGCTGGGTTT AGGCCCCTGG TGTCTGAGGG CCCAGGCCTT GCCAGCCTCT
3251 GCTGCTCCTG CTCCTGGGTT TGAAGATGCA GGCCGATCGC CAGCTCCGTG
3301 GCAGCGGTCA CTAAGGACAG CCTGACTGTG CCATCTTGGG GCCTCAGGCG
3351 GGGCTCCGGA GATAGAAGAC AGGTCGCCGG AGGCTCCCCC TCCTCTCCTC
3401 TCCCTCTGCG AGATGCTCCC TGGGCGCTAC CCTGCAGGGT GCCAGGCAGG
3451 AGTGGTCTCA GAACGTGCGC TTCTGATTAT TTTACTGGGG TCCATTGTCC
3501 AGATTTTCTT TTGATTGTAA AATATATTTT TACTTTTATG TCTTCTAATT
3551 TAATAAATGA TCCATATAAA AATAGAGAAA TAAAGTCCTT TAAGGGAAGG
3601 TTTAAAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 112 bp to 933 bp; peptide length: 274
 Category: similarity to unknown protein
 Classification: no clue

```

1 MATTVSTORG PVYIGELPQD FLRITPTQQQ RQVQLDAQAA QQLQYGGAVG
51 TVGRLNITVV QAKLAKNYGM TRMDPYCRLR LGYAVYETPT AHNGAKNPRW
101 NKVIHCTVPP GVDSFYLEIF DERAFSMDDR IAWTHITPE SLROGKVEDK
151 WYSLSGRQGD DKEGMINLVM SYALLPAAMV MPPQPVLMP TVYQQGVGVY
201 PITGMPAVCS PGMVPVALPP AAVNAQPRCS EEDLKAIQDM FPNMDQEVIR
251 SVLEAQRGNK DAAINSLLOM GEEP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2d15, frame 1

TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2,
 N = 1, Score = 385, P = 1.1e-35

>TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2
 Length = 457

HSPs:

Score = 385 (57.8 bits), Expect = 1.1e-35, P = 1.1e-35
 Identities = 77/182 (42%), Positives = 118/182 (64%)

```

Query:      4 TVSTQRGPVYIGELPQDFLRIT-PTQQQRQVQLDAQAAQQLQYGGAVGTVGRLNITVVQA 62
              TV+ +R V +GELP FLR+ P QQ + ++ Q + + + T GRL++T+++A
Sbjct:      5 TVAERRQVLVGLPFPFLRLAVPIQTAEPEI-VQP-RMVSFVPP-NTRGRLSVTILEA 61

Query:      63 KLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIFDE 122
              L KNYG+ RMDPYCR+R+G ++T A N + P WN+ ++ +P V+S Y++IFDE
Sbjct:      62 NLVKNYGLVRMDPYCRVRVGNVEFDTNVAANAGRAPTWNRTLNAYLPMNVESIYIIFDE 121

Query:      123 RAFSMDRIAWTHITIPESLRQGGKVEDKWYSLSGRQDDKEGMINLVMSYAL--LPAAMV 180
              +AF D+ IAW HI +P ++ G D+++ LSG+QG+ KEGMI+L S+A LP
Sbjct:      122 KAFGPDEVIWAHINLPLAIFNGDNIDEYFQLSGQQGEGKEGMIHLHFSFAPIDLPLQA 181

```


Score = 92 (13.8 bits), Expect = 1.8e-01, P = 1.7e-01
Identities = 26/68 (38%), Positives = 38/68 (55%)

```
Query:   250 RSVLEAQR 257
          + +LE +R
Sbjct:   216 KCILEERR 223
```

Pedant information for DKFZphtes3_2d15, frame 1

Report for DKFZphtes3_2d15.1

```

SEQ      MATTVTSTQRGPVYIGELPQDFLRITPTQQQRQVQLDAQAAQQLQYGGAVGTVGRNLNITVV
SEG      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccceeeecccccceeeeccccchhhhhhhhhhhhhhhhhhhccccceeeecceeh

SEQ      QAKLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIF
SEG      .....
PRD      hhhhhhhhhcccccccchhhhhheeeeeeccccccccccccccccceeeecccccccccceeeec

SEQ      DERAFSMDDRIAWTHITIPESLRQKGVEDKWYLSGRQGDKEGMINLVMSYALLPAAMV
SEG      .....xxxxxxxxxxxxx.....
PRD      cccccccccceeeecccccccccccccccccceeeecccccccccccccceeeehhhhhhhhhhhc

SEQ      MPPQPVLVLMPTVYQQGVGYVPTITGMFAVCSPGMVFPVALPPAAVNAQPRCSEEDLKAIQDM
SEG      xxxxxxxxxxxx.....xxxxxxxxxxxxx.....
PRD      cccccceeeeeeccccccccccccccccceeeccccccccccccceeeeccccchhhhhhhhhc

SEQ      FPNMDQEVIRSVLEAQRGNKDAAINSLLQMGEFP
SEG      .....
PRD      ccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhccc

```

(No Prosite data available for DKFZphtes3_2d15.1)

Pfam for DKFZphtes3_2d15.1

HMM_NAME	C2 domain		
HMM	*LtvRIeARNLWkMDmNgfSDPYVKVdMdPdpkDtkKWKtKiWNNGLN		
		L+++++A+ + + M+ DPY++ + + + +T T +N N	
Query	55	LNITVVQAKLAKNYGMT-RMDPYCRLRLGYAVY-----ETPTAHNGAKN	97
HMM	FVWNEEeFvFedIPyPdlqrkMLRFaVWDWDRFSRBDFIGHCi*		
		P+WN + +P + + + +D+ FS +D I + +	
Query	98	PRWN-KVIHCT-VPPGVDSF---YLEIFDERAFSMDRIAETH	135

DKFZphtes3_2e12

group: Transcription Factors

DKFZphtes3_2e12 encodes a novel 849 amino acid protein with similarity to Zinc finger proteins.

The new protein is a putative transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein, which is only found in cytochrom C related proteins.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to finger proteins

complete cDNA, complete cds, 5 EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3205 bp

Poly A stretch at pos. 3192, polyadenylation signal at pos. 3171

```

1  GGCACGGCCG GGTCTGGCT GGCCAAACGA GGCTCGCGGA AGCAGCAGCC
51  GCCGCTGAC CGCAGCTGGA TTTGAAGAT TGATCCAAGG GACTGTATTA
101 ATTTTCAGGAA TTGATTGAA AGACACTGGC TCTGCCACTT AACAGCCATG
151 TAACCTTGGA TATGGAAGAA AGTAGCAGTG TTGCCATGTT GGTGCCAGAT
201 ATGGGGAAC AGGAAGCTAT ACTGACTGCT GAAAGTATCA TCAGTCCTTC
251 ATTGGAATT GATGAACAAA GAAAACTAA ACCAGATCCA TTAATCCATG
301 TTATCCAGAA GTTAAGCAAG ATAGAAAAAT GAAAAGTCAC AAAATGTCT
351 TTTAATTGGG AAGAAACGCC CACGTTCAAG TGCTGCAACA CACTCTCTTG
401 AAACCAAGA ACTTTGTGAG ATTCGGGCTA AAGTAATCCA GTCACCTGCT
451 GCTGATACTA GAAGGGCTGA GATGTCACAA ACAAATTTA CCCCTGACAC
501 TCTTGCCCGA AATGAAGGGA AGGCTATGTC TTATCAGTGT AGCCTTTGTA
551 AGTTTCTATC ATCATCCTTT TCCGTGTAA AAGATCATAT TAAGCAACAT
601 GGTACAGCAA ATGAAGTGAT ACTGATGTGC TCAGAGTGCC ATATTACATC
651 TAGAAGCCAG GAGGAACCTG AAGCCACGTT GGTGAATGAC CATGACAATG
701 ATGCCAATAT CCACACCCAA TCCAAAGCCC AACAGTGCCT AAGCCCCCTC
751 AGCTCTTTGT GTCGGAAAAC CACAGAAAGA AATGAAACCA TTCCAGATAT
801 CCCAGTAAGT GTGGACAATC TACAGACTCA TACTGTCCAA ACTGCATCTG
851 TGGCAGAAAT GGGTAGGAGG AAATGGTATG CATACGAACA GTACGGCATG
901 TATCGATGCT TGTTTGTAG TTATACTTGT GGCCAGCAGA GAATGTTGAA
951 AACACACGCT TGGAAACATG CTGGGGAGGT TGATTGCTCC TATCCAATCT
1001 TTGAAAATGA AAATGAACCC CTAGGCTGTC TGGATTCTTC AGCAGCTGCT
1051 GCGCCTGGTG GGGTCGATGC AGTCGTCAAT GCTATTGGAG AGAGTGAAC
1101 GAGTATCCAC AATGGGCCAT CAGTGCAAGT GCAGATTGTC AGCTCAGAAC
1151 AGTTATCATC TTCATCTCCT TTAGAACAGA GTGCAGAAAG AGGAGTACAC
1201 CTAAGTCAGT CAGTTACCCCT GGACCCCAAT GAGGAAGAAA TGCTAGAAGT
1251 GATTTCTGAT GCAGAGGAGA ATCTGATTCC TGATAGCCTG CTTACATCAG
1301 CACAGAAAAT CATCAGCAGC AGCCCCAATA AAAAAGGGCA TGTTAACGTG
1351 ATAGTGGAGC GATTGCCAAG TGCTGAAGAA ACCCTTTCAC AGAAGCGCTT
1401 CCTCATGAAC ACTGAAATGG AAGAAGGGAA GGACCTGAGC CTGACAGAAG
1451 CTCAGATTGG GCGCGAAGGA ATGGATGATG TTTATCGTGC TGATAAATGT
1501 ACTGTTGATA TTGGGGGATT GATCATAGGC TGGAGCAGTT CAGAGAAAAA
1551 AGACGAGTTA ATGAATAAAG GCCTGGCTAC TGATGAGAAAT GCCCCACCAG
1601 GCGCGAGAAG GACAAATTCT GAGTCTCTTC GATTACACTC ATTAGCTGCA
1651 GAAGCCCTTG TCACAATGCC TATAAGAGCT GCAGAGTTGA CAAGAGCCAA
1701 CCTGGGGCAC TATGGAGATA TAAACCTTTT AGATCCAGAT ACTAGTCAAA
1751 GGCAAGTAGA TAGTACATTG GCAGCGTACT CAAAATGAT GTCGCCACTT
1801 AAAAATCTCT CAGATGGATT AACTAGTCTT AACCAGAGCA ACTCCACCTT
1851 GGTAGCACTC CCAGAGGGTA GGCAGGAATT GTCAGATGGG CAGGTTAAGA
1901 CAGGCATCAG CATGTCCTTA CTCACCGTCA TTGAAAAATT GAGAGAAAGG
1951 ACAGACCAAA ACGCTTCAGA CGATGACATT TTGAAAGAGT TGCAGGACAA
2001 CGCCCAAGTG CAACCAACA GCGATACAAG TTTGTCCGGA AACAAATGTT
2051 TGGGAATACAT CCCGAATGCT GAACGACCCCT ACCGTGCGG CCTGTGTAC
2101 TACACAAGTG GCAACAAGGG CTACATCAAG CAGCACTTAC GAGTCCATCG
2151 ACAGAGACAG CCTTATCAGT GTCCTATCTG CGAGCACATA GCGGACACA
2201 GCAAAGATTG GGAGAGTCAC ATGATCCACC ACTGTAAGAC AAGAATATAC
2251 CAGTGCAAGC AGTGTGAAGA ATCTTCCAT TATAAGAGTC AATTGAGGAA
2301 CCATGAGAGC GAACAGCACA GTCTTCCAGA TACCTTGTC AATAGCAACTT
2351 CTAATGAGCC AAGAATTTCC AGTGATACAG CTGATGGAAT ATGTGTCCAG
2401 GAAGGGAATA AGTCTTCAGT CCAGAAACAA TATAGATGTG ATGTGTGTGA
2451 TTATACAAGT ACAACATATG TTGGTGTGAG AAACCACAGG CGAATCCATA
2501 ACTCTGATAA GCCGTACAGA TGCTCTCTGT GTGGGTATGT GTGTAGCCAT
2551 CCTCTCTCTT TGAAGTCTCA TATGTGGAAT CATGCAAGTG ACCAAAATTA

```



```

2601 CAACTACGAA CAAGTAAACA AGGCTATTAA CGACGCGATT TCACAAAGTG
2651 GCAGAGTTCT GGGGAAATCC CCTGGAAAGA CTCATTAAA GAGCAGTGAA
2701 GAGAGTGCAG ATCCCGTCAC TGGAGTTTCG GAAATGCAG TGTCTCTTC
2751 AGAACTGATG TCCCAGACTC CCAGTGAAGT TCTGGGTACC AACGAGAATG
2801 AGAACTGAG CCCTACAAGT AATACCTCAT ATAGTTTGA AAAAACTCC
2851 AGTCTGGCCC CTCCTAGCAT GGAGTACTGC GTTTACTCT TCTGCTGTTG
2901 TATTTGTTGG TTTGAATCAA CCAGCAAAGA AAACCTCTG GATCATATGA
2951 AAGAGCACGA GGGTGAAATT GTAAACATCA TCCTGAATAA GGACCACAAT
3001 ACAGCTCTAA ACACAAATTA GGTGGAATAA TGACTCGAGC AGGAAAGCAG
3051 TAGAAGAGGA TTCCTTCACC ACAGTTTCAC CTTTACGCTG TCAGACAACT
3101 TCCTGCCACA GAAGAAGTCG TTGATGTGAT TTTTGAGGAA ATGACAGATG
3151 TGACTTTTGA ACCAACTTG TAATAAAGG AATTCCAAAT GGAAAAAATA
3201 AAAAA

```

BLAST Results

No BLAST result

Medline entries

90301500:
Cloning and sequencing of a zinc finger cDNA expressed in mouse testis.

92310982:
Zfp-37, a new murine zinc finger encoding gene, is expressed in a developmentally regulated pattern in the male germ line.

Peptide information for frame 1

ORF from 472 bp to 3018 bp; peptide length: 849
Category: similarity to known protein

```

1 MSQTNFTPTD LAQNEGKAMS YQCSLCKFLS SSFVSLKDHI KQHGQONEVI
51 LMCSECHITS RSQEELEAHV VNDHDNDANI HTQSKAQOCV SPSSSLCRKT
101 TERNETIPDI PVSVDNLQTH TVQTASVAEM GRRKWAYAYE YGMYRCLFCS
151 YTCGQORMLK THAWKHAGEV DCSYPIFENE NEPLGLLDSS AAAAPGGVDA
201 VVIAIGESL SIHNGPSVQV QICSSEQLSS SSPLEQSAER GVHLSQSVTL
251 DPNEEEMLEV ISDAEENLIP DSLTSAQKI ISSSPNKKGH VNVIVERLPS
301 AEETLSQKRF LMNTEMEEGK DLSLTEAQIG REGMDDVYRA DKCTVDIGGL
351 IIGWSSSEKK DELMNKGLAT DENAPPGRRR TNSESLRLHS LAEEALVTMP
401 IRAAELTRAN LGHYGDINLL DPOTSQRQVD STLAAYSKMM SPLKNSSDGL
451 TSLNQSNTSL VALPEGRQEL SDGQVKTGIS MSLTIVIEKL RERTDQNASD
501 DDILKELQDN AQCQPNSDTS LSGNNVVEYI PNAERPYRCR LCHYTSGNKG
551 YIKQHLRVHR QRQPYQCPIC EHIADNSKDL ESHMIHCKT RIYQCKQCEE
601 SFHYKSQLRN HEREQHSLPD TLSIATSNEP RISSDTADGK CVQEGNKSSV
651 QKQYRCDVCD YTSTTYVGVR NHRRIHNSDK PYRCSLCGYV CSHPPSLKSH
701 MWKHASDQNY NYEQVNKAIN DAISQSGRVL GKSPGKTQLK SSEESADPVT
751 GSSENAVSSS ELMSQTPSEV LGTNEKELK PTSNTSYSLE KISSLAPPSM
801 EYCVLLFCCC ICGFESTSKE NLLDHMKHE GEIVNIILNK DHNTALNTN

```

BLASTP hits

Entry S10245 from database PIR:
finger protein, testis - mouse
Score = 265, P = 8.4e-23, identities = 61/205, positives = 91/205

Entry S22954 from database PIR:
finger protein zfp-37 - mouse
Score = 265, P = 9.1e-22, identities = 61/205, positives = 91/205

Entry AF031657_1 from database TREMBL:
gene: "Zfp94"; product: "zinc-finger protein 94"; Rattus norvegicus
zinc-finger protein 94 (Zfp94) gene, partial cds.
Score = 243, P = 1.6e-21, identities = 57/190, positives = 85/190

Alert BLASTP hits for DKFZphtes3_2e12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2e12, frame 1

Report for DKFZphtes3_2e12.1

[LENGTH] 849
[MW] 94325.42
[pI] 5.47
[HOMOL] PIR:A54661 zinc finger protein ZNF41 - human (fragment) 2e-22
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YJL056c] 3e-09
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YJL056c] 3e-09
[FUNCAT] 04.03.01 trna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT] 04.01.01 rrna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YOR113w] 4e-07
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YGL209w] 2e-04
[FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YNL027w] 2e-04
[FUNCAT] 11.01 stress response [S. cerevisiae, YMR037c] 3e-04
[BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins
[SCOP] dlmeyg_9.6.1.1.1 a designed zinc finger protein [syntheti 8e-06
[PIRKW] nucleus 8e-18
[PIRKW] RNA binding 5e-13
[PIRKW] duplication 7e-13
[PIRKW] tandem repeat 1e-21
[PIRKW] spermatogenesis 6e-16
[PIRKW] zinc 9e-21
[PIRKW] zinc finger 1e-21
[PIRKW] DNA binding 1e-21
[PIRKW] metal binding 3e-15
[PIRKW] phosphoprotein 5e-13
[PIRKW] leucine zipper 1e-13
[PIRKW] alternative splicing 6e-18
[PIRKW] eye lens 2e-16
[PIRKW] oocyte 1e-12
[PIRKW] transcription factor 6e-18
[PIRKW] segmentation 7e-13
[PIRKW] embryo 1e-12
[PIRKW] transcription regulation 2e-19
[PIRKW] homeobox 2e-08
[SUPFAM] POZ domain homology 7e-15
[SUPFAM] transcription factor Krueppel 7e-13
[SUPFAM] zinc finger protein ZFP-36 1e-21
[SUPFAM] homeobox homology 2e-08
[SUPFAM] unassigned homeobox proteins 2e-08
[PROSITE] CYTOCHROME_C 1
[PROSITE] MYRISTYL 10
[PROSITE] ZINC_FINGER_C2H2 3
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 18
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 7
[PFAM] Zinc finger, C2H2 type
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 5.65 %

SEQ MSQTNFTPDTLAQNEGKAMSYQCSLCKFLSSSFVLKDHKQHQQQNEVILMCSECHITS
SEGXXXXXXXXXXXXXXXXXXXXX.....
lmeyF

SEQ RSQEELEAHVVNDHDNDANIHTQSKAQCVSPSSSLCRKTTERNETIPDIPVSDNLQTH
SEG
lmeyF

SEQ TVQTASVAEMGRRKWAYEYQYGMRYCLFCSYTCGQQRMLKTHAWKHAGEVDCSYPIFENE
SEG
lmeyF

SEQ NEPLGLLDSSAAAAAPGGVDAVVIAIGSELSIHNGPSVQVQICSSEQLSSSSPLEQSAER
SEGXXXXXXXXXXXXXXXXXXXXX.....
lmeyF

SEQ GVHLSQSVTLDPNEEEMLEVISDAEENLIPDSLLTSAQKISSSPNKKGHVNVIVERLPS
SEG
lmeyF


```

SEQ      AEETLSQKRFMLNTEMEEGKDLSTEALQIGREGMDDVYRADKCTVDIGGLIIGWSSEKK
SEG      .....
lmeYF    .....

SEQ      DELMNKGLATDENAPPGRRTNSESRLHSLAAEALVTMPIRAAELTRANLGHYGDINLL
SEG      .....
lmeYF    .....

SEQ      DPDTSQRQVDSTLAAYSKMMSPLKNSSDGLTSLNQSNSTLVALPEGRQELSDGQVKTGIS
SEG      .....
lmeYF    .....

SEQ      MSLLTVIEKLRERTDQNASDDILKELQDNAQCQPNSDTSLSGNNVVEYIPNAERPYRCR
SEG      .....
lmeYF    .....TTTEETT

SEQ      LCHYTSNGKGYIKQHLRVHRQRPYQCPICEHIADNSKDLESHMIHCKTRIYQCKQCEE
SEG      .....
lmeYF    TTTCEETTHHHHHHHHHHTTCCEEETTTTEECCHHHHHHHHHHHCCCCCEEETTTTE

SEQ      SFHYKSQLRNHEREQHSLPDTLSIATSNEPRISSDTADGKCVQEGNKSSVQKQYRCDVCD
SEG      .....
lmeYF    EECCHHHHHHHHHHHHC.....

SEQ      YTSTTYGVRNHRRIHNSDKPYRCSLCGYVCSHPPSLKSHMWKHASDQNYNYEQVNKAIN
SEG      .....
lmeYF    .....

SEQ      DAISQSGRVLGKSPGKTQLKSSEESADPVTGSSSENAVSSSELMSQTPSEVLGTNENEKLS
SEG      .....
lmeYF    .....

SEQ      PTSNTSYSLEKISSLAPPSMEYCVLLFCCCICGFESTSKENLLDHMKHEGEIVNIILNK
SEG      .....
lmeYF    .....

SEQ      DHNTALNTN
SEG      .....
lmeYF    .....

```

Prosites for DKFZphtes3_2el2.1

PS000001	104->108	ASN_GLYCOSYLATION	PDOC00001
PS000001	445->449	ASN_GLYCOSYLATION	PDOC00001
PS000001	454->458	ASN_GLYCOSYLATION	PDOC00001
PS000001	457->461	ASN_GLYCOSYLATION	PDOC00001
PS000001	497->501	ASN_GLYCOSYLATION	PDOC00001
PS000001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS000001	784->788	ASN_GLYCOSYLATION	PDOC00001
PS000004	98->102	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS000005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS000005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS000005	357->360	PKC_PHOSPHO_SITE	PDOC00005
PS000005	385->388	PKC_PHOSPHO_SITE	PDOC00005
PS000005	425->428	PKC_PHOSPHO_SITE	PDOC00005
PS000005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS000005	696->699	PKC_PHOSPHO_SITE	PDOC00005
PS000005	726->729	PKC_PHOSPHO_SITE	PDOC00005
PS000005	817->820	PKC_PHOSPHO_SITE	PDOC00005
PS000006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS000006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS000006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS000006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS000006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS000006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS000006	314->318	CK2_PHOSPHO_SITE	PDOC00006
PS000006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS000006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS000006	381->385	CK2_PHOSPHO_SITE	PDOC00006
PS000006	485->489	CK2_PHOSPHO_SITE	PDOC00006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS000006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS000006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS000006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS000006	758->762	CK2_PHOSPHO_SITE	PDOC00006
PS000006	766->770	CK2_PHOSPHO_SITE	PDOC00006
PS000006	817->821	CK2_PHOSPHO_SITE	PDOC00006

PS00007	331->339	TYR_PHOSPHO_SITE	PDOC00007
PS00007	703->711	TYR_PHOSPHO_SITE	PDOC00007
PS00007	596->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	142->148	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	241->247	MYRISTYL	PDOC00008
PS00008	349->355	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	478->484	MYRISTYL	PDOC00008
PS00008	645->651	MYRISTYL	PDOC00008
PS00008	751->757	MYRISTYL	PDOC00008
PS00008	772->778	MYRISTYL	PDOC00008
PS00009	130->134	AMIDATION	PDOC00009
PS00009	376->380	AMIDATION	PDOC00009
PS00028	146->167	ZINC_FINGER_C2H2	PDOC00028
PS00028	684->705	ZINC_FINGER_C2H2	PDOC00028
PS00028	595->617	ZINC_FINGER_C2H2	PDOC00028
PS00190	53->59	CYTCHROME_C	PDOC00169

Pfam for DKFzphtes3_2e12.1

HMM_NAME Zinc finger, C2H2 type

HMM *CpWPDCgKtFrrwsNlRrHMR.T.H*
 C++ C+ T R+++L++H H
 Query 53 CSE--CHITSRSQEELEAHVVN-DH 74

23.25 (bits) f: 539 t: 559 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

Query *CpWPDCgKtFrrwsNlRrHMRTH*
 C C++T ++ ++H+R+H
 dkfzphtes3 539 CRL--CHYTSGNKGyIKQHRLRVH 559

Query f: 567 t: 587 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

HMM *CpWPDCgKtFrrwsNlRrHMRTH*
 CP+ C+ ++ +L+ HM+ H
 Query 567 CPI--CEHIADNSKDLESHMIHH 587

33.47 (bits) f: 595 t: 616 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

Query *CpWPDCgKtFrrwsNlRrHMR.T.H*
 C+ C+++F ++S+LR+H R H
 dkfzphtes3 595 CKQ--CEESFHYKSQLRNHERE-QH 616

Query f: 656 t: 676 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

HMM *CpWPDCgKtFrrwsNlRrHMRTH*
 C++ C++T ++ R+H+R+H
 Query 656 CDV--CDYTSTTYVGVNRHRIH 676

24.53 (bits) f: 684 t: 704 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

Query *CpWPDCgKtFrrwsNlRrHMRTH*
 C+ CG++ +++ +L+ HM H
 dkfzphtes3 684 CSL--CGYVCSHPPSLKSHMWKH 704

Query f: 809 t: 829 Target: dkfzphtes3_2e12.1 similarity to finger proteins
 Alignment to HMM consensus:

HMM *CpWPDCgKtFrrwsNlRrHMRTH*
 C+ CG ++++NL HM+ H
 Query 809 CCI--CGFESTSKENLLDHMKH 829

DKFZphtes3_2f14

group: testes derived

DKFZphtes3_2f14 encodes a novel 129 amino acid protein with very weak similarity to human omega protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to omega protein

complete cDNA, complete cds, 1 EST hit

Sequenced by EMBL

Locus: unknown

Insert length: 2353 bp

Poly A stretch at pos. 2341, no polyadenylation signal found

```
1 GCAGATTCTC CAGGCCAGC ATCTGCCTCA CCGTGGCCCC CCACAAGCCA
51 AGCGCCTGCC TTTCAGCAGC CTCTACACAC CCAGCTCCTG CCACCCAATG
101 GCTCTTTAGG CCAAGCTCAT ACCTCAGCAT GATTTTCCCA GGCCCAACTT
151 TTGTCTCATG GCAACCTTCC CTGGCCAAGT TTCCACCTAT TTCCTGGCAG
201 CCTGGACAGG CCCAGGTCCT GCCACACACT GGCTCTCTTA CGCCAGCTC
251 ATGCCTCACA GTGGCCTCTC CAGGCCAGC TCCTGTCCCG GGACATCATC
301 TCCAGGCCCA AAACCTTCTC AAGTCGGCCT CTCCAGGCCC AGTTGCTGCC
351 TCCCGGCATT CTCTCCAGGC CTAGCTCTTC CTCCTGGCTG TATCTACAAG
401 ACCAACTCCT GCCTCACAAAC AACCTTTTAT GGCTCAGCTC CTGCCCAACT
451 ACTGCCCGGCC TTTGTAGGCC CAAAACCTCC TCAAGTCAAG CTCTTTAGGC
501 CCACCTTCTG CCTTGCAGTG GCCTGTACAG ACCCAGCTCT GGCTTGAGAA
551 CAGCCTCTGC AGGCCCTGCT CTGCTCTT AGCTCCCTCT CCAGGCCCAT
601 CTCTTGCCCTC ACAGTGGCTT CCGTGGGCCA AGTTCCCGCC TGCCTCCAG
651 CAGCCTCAAC AGGCCTAGCT CCTCCCTCAC AATGGCTTGT TTAGGTCCAG
701 TTGATGCCTC TGGCAACCTG TCCAGGCCCA GCTCCTGCCT CACACTGGCC
751 TCTCTAGGCC GAGGTCCTTT CTCATACTGG CCTGTTTAGG CCCAGCTCAT
801 TCCTCTTGTC ATCTCTCCAG GCCCAGCTTT TGCTGTGTTG TGGCCTCTAC
851 CTCACAGTGC ACCTTCCAGT CCCACCTCTT GCCTCACCAT GGCCTCTCTT
901 GACCAGGTTT CTGCCTTTTG GCAGCTCTTA CAGGCCTAGC TGCTGCCTCC
951 CAATGGCCTT TGTAGGCCAC GCTCATGCC TACTGTGGCC TTTCAGGCC
1001 TAGCTTTCCG TTTTGGCCA CTCCAGGCCC AGAACTTCCC CCAGTCAGCC
1051 TCTCCAGGCC CAGCTCTTCC TCCAGCAAC CTCTGCAGGC CCAAATCATC
1101 CTCAAATTGG CCTCTCTTT CCCAGCTCTT GCCTCCTGGT GGCCTCTGAA
1151 GACCCAAATC TCTCTCCAGT TGGTTTTTCC AGGCCAGCT CCTGCCTTTT
1201 GGTGGCCTCT CCAGGTGCAA AACTTCTTCC CATCAGCCTG TCCAGGCCCA
1251 GCTCATGCC TTTGGTGGCC TTCTCAGGCC CTGCTTTTGA CTTGGTGGCC
1301 TCTTCAGGCC CAGAACTTGA ACTCAAGTCA GCCTCTCCAG GCCCAGCTCC
1351 TGCTTCTTFA AGTCTGTAC AGGCCAGGCC TCTACCTCAC AGCGGACTCT
1401 CCACACCCAG CTCTTGCCCT ACTGTAGCCT CCCAGTCCA AAACCTCTGC
1451 CTTTTGGCAG CTTCGACAAG CCCAGCTCCT GCCTTTCAAT GACCTCTTTA
1501 GGCCCGCTC ATTCCTTACA ACGGCCCTTC CAGGCCAGT TTTTCCCTTT
1551 TGGCGGCCTC TCCAGGCCCA GAACTTCTC AAGTCGGCCT CTTTAGGCCC
1601 AGTTGCTGCC TCCTGGCATC CTCTGCAGGC CGAGCTCTTC CTCCTGCTG
1651 TGTCTACAGG CCCAACTCCT GCCTCACAAC AACCTCTTGG GACTCAGCTT
1701 CTGCCAGCT CCTGGTGGCC TTTGTAGGCT CAAAATTTTC TCAAATCAAG
1751 CTCTCCAGGC CTACTGTCAG CCTCGTGGCA GCCTAAACAG GCCCAGCTCC
1801 TGCTTGACAA TGGCCTCTCC AGGCTTTTCT CTTGCCTCGC AGCAGGCTTT
1851 CCAGGCCAG CTCTTGCCCT ATGGTGGCCT TCCCCGGCCA TGTTCCTATC
1901 TGACTTCTGG CAGCCTCAAC CGGCCAGCT TCTGCCTCAC ACTGGCCTCT
1951 CTAGGCCAG CTCTTTTTC ACAGTGGCCT CACTACGCCC ATCTCTTACC
2001 TCAGATCTGC CTCCCAAGAC CCAGCTCCTG TCTCATGGTG GTCTCTCTTA
2051 CACCAGCTCC TGCCTCACA TGGCCTCGTC TGGCCCATCT TCTGCCTCAC
2101 AGTGGCCACT CAAGGCCCAT CTTTGGCCTC ATGGTAGCCT CTTCTGGTTT
2151 TGCTCTTGGC TCACAGTTGC CTCTTCCAGA TCCAGCTTTA AGCCTTTGAT
2201 GGTCAACAGC ATCAAGGAGC CTAAGGCTTC CCTGGACTCT CATTTGTTC
2251 CTTTACAGCA GAGTGCCTTA GCAAAACTG TCTCTTAACC TTGAGAGTGG
2301 ATTTCTGACA AATCGATAGT AAATTCCTGC TGTGTGGTTT CAAAAAATAA
2351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 158 bp to 544 bp; peptide length: 129
 Category: similarity to known protein

1 MATFPGQVST YFLAAWTGPG PATHWPLYAQ LMPHSGLSRP SSCPGTSSPG
 51 PKLPQVGLSR PSCCLPAFSP GLALPPGCIY KTNCLTTTF YGSAPAQLLP
 101 AFVGPQLPQV KLFRTFCLA VACTDPALA

BLASTP hits

Entry I70697 from database PIR:
 omega protein - human (fragment)
 Score = 79, P = 2.8e-03, identities = 32/94, positives = 38/94

Alert BLASTP hits for DKFZphtes3_2f14, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2f14, frame 2

Report for DKFZphtes3_2f14.2

[LENGTH] 129
 [MW] 13421.76
 [pI] 9.14
 [PROSITE] MYRISTYL 2
 [KW] Irregular
 [KW] LOW_COMPLEXITY 10.85 %

SEQ MATFPGQVSTYFLAAWTGPGPATHWPLYAQLMPHSGLSRPSSCPGTSSPGPKLPQVGLSR
 SEGxxxxxxxxxxxxxxxx.....
 PRD cccccccceehhhhhccc
 SEQ PSCCLPAFSPGLALPPGCIYKTNCLTTTFYGSAPAQLLP AFVGPQLPQVKLFRTFCLA
 SEG
 PRD ccccccccccccccccccccccccccecccccccccccccccccccccccccccccccccc
 SEQ VACTDPALA
 SEG
 PRD ccccccccc

Prosite for DKFZphtes3_2f14.2

PS00008	6->12	MYRISTYL	PDOC00008
PS00008	92->98	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2f14.2)

DKFZphtes3_2g7

group: testes derived

DKFZphtes3_2g7 encodes a novel 359 amino acid protein with similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

complete cDNA, complete cds, 6 EST hits (5 hits are out of a testis library)

Sequenced by EMBL

Locus: unknown

Insert length: 1613 bp

Poly A stretch at pos. 1595, polyadenylation signal at pos. 1557

```
1 GCCACACAGG CTCCTTGGAG TAAGAGTGTG AGAACTGGA TGAAGACAGC
51 TGTATTCTTT TGAAGCGTTC CGAGATTGGT CTGTCTCTAC CAACTAAAAA
101 CTTCTAGCTT AAGTGCAGAG ATTTAAGGAG ATCAACAAAA ACTCAGTCTA
151 GACATATTAT GAGGCTGGGA GGGTATCAAC AGACTTGAGT TCTGTGTCAGC
201 AAGATCACCT GCTTTTAATA TTGTCTCTAG GGTCTGAGCA CATCTGGAAG
251 TGAGGTCAAT CAAGTTAGAC CCCAAAAACT TTTGTGACAA CAGTGAAGAG
301 GGGAAAAATA ACACACCACA AACATGAACC TCAACCCCCC GACATCTGCT
351 CTTGAGATCG AGGGCAAAGG CAGCCATATT ATGGCTAGAA ATGTAAGCTG
401 CTTTCTAGTC AGGCACACCC CTCATCCCAG AAGAGTCTGC CACATCAAAG
451 GCTTGAATAA CATTCCAATC TGTACTGTGA ATGATGATGA GAATGCATTT
501 GGAACATTGT GGAAGTTTGG CCAGTCTAAC TACTTAGAGA AGAACAGGAT
551 ACCATTGGCC AATTGCAGTT ACCCCCCGAG CACTGCAGTC CAGAAGAGCC
601 CTGTAAGAGG AATGTCGCCA GCCCCAAACG GTGCCAAAGT GCCTCCACGG
651 CCTCATCTCG AGCCCAGTAG AAAAATTAAA GAGTGCTTCA AAACCTCCAG
701 TGAGAAATCCC TTAGTAATTA AAAAGGAAGA AATTAAGGCC AAAAGACCAC
751 CATCACTCC AAAGGCATGC TCTACTCCTG GCTCCTGTTC TTCAGGGATG
801 ACAAGTACCA AGAATGATGT GAAAGCAAAC ACCATTGCA TACCAAACTA
851 TCTGGATCAG GAAATAAAAA TCCTGGCAAA GCTCTGTAGC ATTTTGCATA
901 CTGATTCTCT GGCAGAAAGT TTACAGTGGC TGCTTCATGC AACTTCAAAA
951 GAAAAAGAGT GGGTCTCAGC TTTGATTTCAT TCTGAGCTTG CCGAGATAAA
1001 CCTGTAACT CATCACAGAA GAAACACCTC AATGGAACCA GCAGCAGAGA
1051 CTGGGAAGCC ACCCACAGTT AAATCACCAC CCACAGTTAA ATTGCCCCCA
1101 AATTTTACTG CAAAATCAAA AGTGCTGACC AGAGATACAG AAGGGGATCA
1151 ACCAACCAGA GTGTCAAGTC AAGGATCTGA AGAAAAACAAG GAAGTACCAA
1201 AAGAGGCTGA GCACAAGCCT CCACTACTTA TAAGAAGAAA TAATATGAAA
1251 ATACCTGTTG CAGAATATTT CAGCAAACCA AATTCTCCTC CCAGGCCTAA
1301 CACTCAGGAG AGTGGATCAG CAAAACCAAG GTGAGCAAGG AGTATACAAG
1351 AATACAACCT CTGTCCCCAA AGAGCATGTT ATCCTTCAAC ACACCGGAGG
1401 TAGAAGTTCT AGACTGGGTG AATTCTTTCA TGAATATGAG CTTACATTT
1451 ACATCATCAA ATTATTTTTC AAATGAATAT TTTTGGTATT GAGGAATCAA
1501 GTGGTCTCTT TTATGGTGGC ACATGTAAAT CTAATAATAC CTGTATGTAA
1551 TGCTACAAAT AAATATTACT GGAAATGATA TTTCCATTG TAGTTAAAAA
1601 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 324 bp to 1400 bp; peptide length: 359
Category: similarity to known protein


```

1  MNLNPTPTAL QIEGKGSHIM ARNVSCFLVR HTPHPRRVCH IKGLNNIPIC
51 TVNDDENAFG TLWEVGQSNY LEKNRIPFAN CSYPPSTAVQ KSPVRGMSPA
101 PNGAKVPPRP HSEPSRKIKE CFKTSSSENPL VIKKEEIKAK RPPSPPKACS
151 TPGSCSSGMT STKNDVKANT ICIPNYLDQE IKILAKLCSI LHTDSLAEVL
201 QWLLHATSKE KEWVSALIHS ELAEINLLTH HRRNTSMEPA AETGKPPTVK
251 SPPTVKLPPN FTAQSKVLTR DTEGDQPTRV SSQGSEENKE VPKEAEHKPP
301 LLIRRNMMKI PVAEYFSKPN SPPRPNTQES GSAKPVARS IQEYNLCPQR
351 ACYPSTHRR

```

BLASTP hits

Entry A43427 from database PIR:
 neurofilament triplet H1 protein - rabbit (fragment)
 Score = 118, P = 5.6e-04, identities = 79/290, positives = 110/290

Entry RNNFH_1 from database TREMBL:
 Rat high molecular weight neurofilament (NF-H) protein mRNA, 3' end.
 Score = 115, P = 9.5e-04, identities = 69/281, positives = 100/281

Entry B43427 from database PIR:
 neurofilament protein H form H2 (repetitive region) - rabbit (fragment)
 Score = 111, P = 1.3e-03, identities = 64/269, positives = 102/269

Alert BLASTP hits for DKFZphtes3_2g7, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2g7, frame 3

Report for DKFZphtes3_2g7.3

```

[LENGTH]      359
[MW]           39725.53
[PI]           9.45
[PROSITE]      MYRISTYL      3
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      9
[PROSITE]      PKC_PHOSPHO_SITE      10
[PROSITE]      ASN_GLYCOSYLATION      4
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      4.18 %

SEQ  MNLNPTPTALQIEGKGSHIMARNVSCFLVRHTPHPRRVCHIKGLNNIPICTVNDDENAFG
SEG  .....
PRD  cccccccccceccccceeecccccccccccccccccccccccccccccccccccccc

SEQ  TLWEVGQSNYLEKNRIPFANCSYPPSTAVQKSPVRGMSAPNGAKVPPRP HSEPSRKIKE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhh

SEQ  CFKTSSSENPLVIKKEEIKAKRPPSPPKACSTPGSCSSGMTSTKNDVKANTICIPNYLDQE
SEG  .....
PRD  hccccccccceehhhhhhhccccccccccccccccccccccccccccccccccccceehhh

SEQ  IKILAKLCSILHTDSLAEVLQWLLHATSKEKEWVSALIHS ELAEINLLTHHRRNTSMEPA
SEG  .....
PRD  hhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc

SEQ  AETGKPPTVKSPPTVKLPPNFTAQSKVLTRDTEGDQPTRVSSQGSEENKEVPKEAEHKPP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LLIRRNMMKIPVAEYFSKPN SPPRPNTQESGSAKPVARS IQEYNLCPQRACYPSTHRR
SEG  .....
PRD  eeeccccccccceccccccccccccccccccccccccchhhhhcccccccccccccccccc

```

Prosites for DKFZphtes3_2g7.3

```

PS00001      23->27  ASN_GLYCOSYLATION      PDOC00001
PS00001      80->84  ASN_GLYCOSYLATION      PDOC00001
PS00001     234->238  ASN_GLYCOSYLATION      PDOC00001

```


PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	262->265	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	356->359	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	207->211	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00006	272->276	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	158->164	MYRISTYL	PDOC00008
PS00008	284->290	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphtes3_2g7.3)

DKFZphtes3_2h1

group: transmembrane protein

DKFZphtes3 2h1 encodes a novel 116 amino acid protein with weak similarity to C. elegans cosmid C13F10.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to C.elegans C13F10.5

TRANSMEMBRANE 1

Sequenced by EMBL

Locus: /map="2"

Insert length: 1156 bp

Poly A stretch at pos. 1143, polyadenylation signal at pos. 1121

```

1  GGCCATCAAA ATAACAAAC CATGTCATT GGAGCAACAA AGCCACTGCG
51  GCCTCCATT GGGCCAAGCT CTGACTGCAA TGATGCCTCT GCGCCGACCC
101  GGGCCTCGCT GTGACTGACA ATGCCGCTGC ATCTTTTCAG CAGTCATTGA
151  TGAGGAAGTA TCTACATCCT CCTTCCCACT ACCAGATTTT GCTTGGAGAA
201  AAGCAGTTTC CTGAAATAAT TCTGTGACGA GCTTCTTCCA CATTAGGACA
251  AAAATGCTGG AAGCGGCTCA GCGCCAGGGC AGCACATCAG AGACACCATG
301  GAACACAGCC ATTCCTCTGC CGTCGTGCTG GGACCACTCT TTCCTGACCA
351  ATATCACCTT CTGGAAGGTT CTCTCTGGT TGGTCCTGCT GGGACTGTTT
401  GTGGAACTGG AATTGGCCCT GGCATATTTT GTCCTGTCTT TGTCTATTG
451  GATGTACGTC GGGACACGAG GCCCTGAAGA GAAGAAAGAG GGAGAGAAGA
501  GCGCCTACTC TGTGTTCAAT CCAGGCTGTG AAGCCATCCA GGGCACCCCTG
551  ACTGCAGAGC AGTTGGAGCG CGAGTTACAG TTGAGACCCC TGGCAGGGAG
601  ATAGGACCCA GCTGTGCTGT CATGCAGCTA ACCTCTGATG TGGTCTTCCT
651  CACCATTTGC TATGGATTG ATTTCAAGTG TATAGGACTA AGGGCAGCTT
701  GCGGGTTAGC TCTGTGACTG CATAGTTTTT CTACCTTCTT TCCTGTATCT
751  TTTGCTGCCA TTTGATCTTT GATAGTTTTG GTGAAACTCT CTAATAATACA
801  TTCACCTGTG GTCCGACGCA ATTTATAAAA ATTATGTAAT CAAGAAGGGA
851  GACCTGTTTG TTTCATTTCT CATCTGTTTG GGAGATGATT TTAGAGCACT
901  AGAAAGGCAC TGGGGAGATT CTCAGCTTAA AACATCCAGC AGTTTGAAGT
951  ATGATTAGGT ACATCAGGGC TGCATTGTCA ATGTTCTCTT TAAGTCTTTT
1001 AACATTTATA GCAATTTTTT TTTTCCCGGA GAGTTTAGGT TGCAAGTTTT
1051 GGGTTTCTTG TTTGTTTTTG TTTTGCTTCC TGCTTTAATT CTTTAATTTT
1101 CAGTCATTAC TGGTATTGAA AAATAAAATA TCTTTAAAC ATCAAAAAAA
1151 AAAAAA

```

BLAST Results

Entry HS313307 from database EMBL:
human STS SHGC-16715.
Score = 1222, P = 1.4e-48, identities = 248/251

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 254 bp to 601 bp; peptide length: 116
Category: similarity to unknown protein

```

1  MLEAAQPGQS TSETPWNTAI PLPSCWDQSF LTNITFLKVL LWLVLGLFV
51  ELEFGLAYFV LSLFYWMYVG TRGPKEKKEG EKSAYSVFNP GCEAIQGTLT
101  AEQLERELQL RPLAGR

```


DKFZphtes3_2h15

group: testes derived

DKFZphtes3_2h15 encodes a novel 855 amino acid protein with very weak similarity to *S. pombe* cdc23.

No informative BLAST results; no predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to cdc23

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4619 bp

Poly A stretch at pos. 4598, polyadenylation signal at pos. 4589

```
1 GAAGGCGTCC CGGCATCGGC CAAGATTCTA CATTGCTCAT CTGGGCATCT
51 GAGCCTCCTT CGAAGTTTCC TGTCACAAC TCTCTCTGA CAGCATGGAT
101 GAGGAGGAAG ACAATCTGTC TCTGCTGACC GCACTGCTGG AAGAAAATGA
151 GTCAGCCTTG GATTGTAATT CAGAAGAAA TAACCTCTTG ACGCGGAAA
201 ATGGCGAGCC CGACGCATTT GATGAGCTCT TTGATGCCGA CGGCGACGGT
251 GAATCTTATA CAGAAAGGCG TGATGATGGA GAAACAGGAG AGACAAGAGA
301 CGAAAAGGAA AATCTGGCCA CTCTCTTTGG AGATATGGAG GACTTAACAG
351 ATGAAGAAGA AGTTCCCGCA TCACAGTCAA CTGAAAATAG GGTCTCCCT
401 GCTCTGCCCC CCAGGCGAGA GAAAACGAAT GAAGAGTTGC AAGAGGAATT
451 AAGGAATTTG CAAGAGCAAA TGAAGGCCTT ACAAGAGCAG CTAAAAGTAA
501 CAACAATTA ACAGACAGCA AGCCAGCCCC GTCTGCAAAA ATCCCTGAG
551 AAGTCTCCCC GGCACCTCT TAAGGAGAGG AGAGTTCAGA GAATTCAGCA
601 GTCAACATGC TTTTCTGCGG AGCTTGATGT CCCTGCGCTA CCAAGAACCA
651 AGAGGGTGGC TCGAACACCA AAGCCTTCAC CTCCAGATCC CAAAAGCTCA
701 TCTTCAAGGA TGACAAGTGC ACCCTCCCAA CCCCTACAGA CGATTCTCG
751 GAACAAACCT AGTGGGATAA CTAGAGGTCA AATTGTGGGG ACCCCAGGAA
801 GTTCTGGGGA AACGACTCAA CCCATCTGTG TGGAAAGCCT CTCTGGTCTG
851 CGGCTCAGGC GGCCTCGAGT ATCCTCCACA GAAATGAACA AGAAAATGAC
901 CGGCCGAAAA CTGATCAGAC TGTCTCAGAT CAAGGAAAAG ATGGCCAGAG
951 AGAAGCTGGA AGAAATAGAT TGGGTGACAT TTGGGGTTAT ATTGAAGAAG
1001 GTTACGCCAC AGAGTGTGAA TAGTGGAAAA ACCTTCAGCA TATGGAACT
1051 GAATGATCTT CGTGACCTGA CACAATGTGT GTCTTGTGTC TTATTGGAG
1101 AAGTTCACAA AGCGCTCTGG AAGACGGAGC AGGGGACTGT CGTAGGGATC
1151 CTCGAATGCCA ACCCATGAA GCCCAAGGAT GGTTCAGAGG AGGTGTGTTT
1201 ATCTATCGAT CATCCTCAGA AGGTCTTAAT TATGGGTGAA GCTCTTGACC
1251 TGGGAACCTG TAAAGCCAAG AAGAAGAATG GAGAGCCGTG CACGCAGACT
1301 GTGAATTTGC GTGACTGTGA GTACTGTGAG TACCATGTCC AGGCTCAGTA
1351 CAAGAAGCTC AGTGCAAGC GTGCGGATCT GCAGTCCACC TTCTCTGGAG
1401 GACGAATTCC AAAGAAGTTT GCCCGCAGAG GCACCAAGCT CAAAGAACGG
1451 CTGTGCCAAG ATGGCTTTTA CTACGGAGGG GTTTCTTCTG CCTCGTATGC
1501 AGCTTCAATT GCAGCAGCTG TGGCTCCTAA GAAGAAGATT CAAACCACCT
1551 TGAGTAATCT GGTGTGTAAG GGCACAAACT TGATCATCCA GGAACACGG
1601 CAAAACTCG GAATACCCCA GAAGAGCCTG TCTTGCTCTG AGGAGTTCAA
1651 GGAACCTGAT GACCTGCCGA CGTGTGGAGC CAGGAACCTA AAACAACATT
1701 TAGCCAAAGC CTCAGCTTCA GGGATTATGG GGAGCCCCAA ACCAGCCATC
1751 AAGTCCATCT CGGCCTCAGC ACTCTTGAAG CAACAGAAGC AGCGGATGTT
1801 GGAGATGAGG AGAAGGAAAT CAGAAGAAAT ACAGAAGCGA TTTCTGCAGA
1851 GCTCAAGTGA AGTTGAGAGC CCAGCTGTGC CATCTTCATC AAGACAGCCC
1901 CCTGCTCAGC CTCCACGGAC AGGATCCGAG TTCCCCAGGC TGGAGGGAGC
1951 CCCGGCCACA ATGACGCCCA AGCTGGGGCG AGGTGCTTTG GAAGGAGATG
2001 ATGTTCTCTT TTATGATGAG TCACCACCAC CAAGACCAAA ACTGAGTGCT
2051 TTAGCAGAAG CAAAAAGTT AGCTGCTATC ACCAAATTAA GGGCAAAAGG
2101 CCAGGTTCCT AAAAAACAA ACCCAAACAG CATTAAAGAAG AAACAAAAGG
2151 ACCCTCAGGA CATCCTGGAG GTGAAGGAAC GTGTAGAAAA AAACACCATG
2201 TTTTCTTCTC AAGCTGAGGA TGAATTGGAG CCTGCCAGGA AAAAAAGGAG
2251 AGAACAACCT GCCTATCTGG AATCTGAGGA ATTTCAAGAA ATCCTAAAAG
2301 CAAAATCAAA ACACACAGGC ATCTGAAAG AGGCCGAGGC TGAGATGCAG
2351 GAGCGCTACT TTGAGCCACT GGTGAAAAAA GAACAAATGG AAGAAAAGAT
2401 GAGAAACATC AGAGAAGTGA AGTGCCGTGT CGTGACATGC AAGACGTGCG
2451 CCTATACCCA CTTCAAGCTG CTGGAGACCT GCGTCAGTGA GCAGCATGAA
2501 TACCACTGGC ATGATGGTGT GAAGAGGTTT TTCAAATGTC CCTGTGAAAA
2551 CAGAAGCATC TCCTTGGACA GACTCCCGAA CAAGCACTGC AGTAACTGTG
2601 CCTCTACAA ATGGGAACGG GACGGAATGC TAAAGGTATG CCATTTGCGT
2651 ACTAATTTTT GACTCCTTTT AGTGACCCAT GCTAATAATG TGGAACCATC
```



```

2701 TCCTATTAAA ATATTTTCAT TTTTCTAGGA AAAGACTGGT CCAAAGATAG
2751 GAGGAGAAAC TCTGTTACCA AGAGGAGAAG AACATGCTAA ATTTCTGAAC
2801 AGCCTTAAAT AACCCGAAC TCAGACATTT TCCCACAGAC TTCCTGGCCT
2851 CCTGTGACTC TGGAAAGCAA AGGATTGGCT GTGTATTGTC CATTGATTCC
2901 TGATTGACGC CGTCAAAAAC AAATGCTTGT TAAGCCCATTA AGCTTTGCCT
2951 GCTTACTTTT TGCCATTGGG TTGGTTTGAT ACCACATTTA ACATTGACAT
3001 TTAAGTGGAA AACCAAGTTA TCATTGTCTT TTCTAAGCTC AGTGTGGATG
3051 ATTGCATTAC TTCATTCACT GAAGTTTTTG CCCAAAATT GGAAGGTAAA
3101 CAGAGAGCTA TGTTTCTGTA TCTTTTGGTT ATAGAGTGT CACTTCTTTA
3151 TCATAACAAA ATTCTAGTGT TTATACGAAC ACCCAGAGGC AAAAGAATTT
3201 GGCTTAATTC TCACCTCAGG TAAGTAGCTT AACTTCTGGG CTTCACTTTT
3251 CTCATCTGTA AAATCAGGAA GATTGGACTA AGTGATCCTG AAATGATTTT
3301 TTTAGCACTG GATTTCTACA AATAATAAAA CTTTCCCATC TAGATAATGA
3351 TGATCACATA GTCTTGATGT ACGGACATTA AAAGCCAGAT TTCTTCATTC
3401 AATTCTGTGA TCTCTGTTTT ACTCTTTGAA ATTGATCAAG CCACTGAATC
3451 ACTTTGCATT TCAGTTTATA TATAGAGAGA GAAAGAAGGC TGTCTGCTCT
3501 TACATTATTG TGGAGCCCTG TGATAGAAAT ATGTAAATC TCATATTATT
3551 TTTTTTTTAA TTTTTTTATT TTTTATGACA GGGTCTCACT ATGTCACCTT
3601 GGCTGGAGTG CAGTAGTGGC ATCGCGGCAC ACTGCAGCCT TGGCTTCCCT
3651 GGGCTCAAGC AGTCCTCCCA CCTCAGTCTC CCAAATAGCT AGGACTACAG
3701 GCGTGCCTGA CCAAGCCAG CTAATTTTGT CATTTTGTG AGAGATGGGG
3751 TTTTGGCATG TTGCTCAGGC TGGTCTCAAA CTCCTGAGCA CTAGCAATCC
3801 ACCCACCTCT GTTTCACAAA AAAAAAATAA AATGAAAGGT CAACCCCTAT
3851 GCAAATTACC ACAGCAAAGG TTTCACTCAG GAGATTCTTC CATCTGGGCA
3901 ACCTGGTTTT CCAATATCA TTTGACCTAA GTGAATGTTG ATACTAGCTA
3951 AAGATTGGGT AAATTGGTTG AATTATTGTA TTGAAGCTTG AGCTGTAGCT
4001 AAAAGTAATT TAGGTTTCCC CTAAGATGTT ATTATGTTAG GGACATAACA
4051 CTTTGGGAG GTTGTGTGG GAGATGGTTG ATTTAGGTTT TCAAAAGCTA
4101 GAAATAAAAT TTACATGCCT TAGATTTCAT AAAATTCTGC TCTAATTGGG
4151 TGGAAAGGTG TGTATCTAAC TTGTGTTCCCT CCTAAGGTTA TGTCTTAATA
4201 ACTATTCTTT TAGGAGTATA CTTCTACTTT ATAGAAGGTT GCTTTTCTTT
4251 TTAATTTTTT CTAACAAAGA AAAGAATAAA GTATTATTA ATAAGAACCA
4301 GAAAGCACTT GAAACTGATG TTTTAAATGG CTCATTTAGG GTAGATTTAT
4351 TTATCTCATT AACTTAAAC AGCTATGTGT ATGAAATAGG TCACAACAGA
4401 ACTTGAACAC CAGGTTGGTG TCTGAGCAAT CCCTTCTTA TGGGAAAAAC
4451 AATGTTCTTG TTTGAACAGA GGGTATCATT GCAGTCAGTA TTCACGTGTA
4501 TATTGTTATA TAAGTTGTAT AATATGCTTG TAAAGGCTGA GGGTGAGCTG
4551 TATCTGGATG CCTTTTACA ATTTGATTTT AACTTTTAAA ATAAATTTAA
4601 AACATAAAAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 95 bp to 2659 bp; peptide length: 855

Category: similarity to known protein

Classification: Cell division

```

1 MDEEEDNLSL LTALLEENES ALDCNSEENN FLTRENGEPD AFDELFDAAG
51 DGESEYTEAD DGETGETRDE KENLATLFGD MEDLTDEEEV PASQSTENRV
101 LPAPAPRREK TNEELQEEEL NLQEQMKALQ EQLKVTTIKQ TASPRLQKS
151 PEKSPRPPLK ERRVQRIQES TCFSAELDVP ALPRTKRVAR TPKPSPPDPK
201 SSSSRMTSAP SQPLQTISRN KPSGITRQOI VGTGPGSSGET TOPICVEAFS
251 GLRLRRPRVS STEMNKKMTG RKLIRLSQIK EKMAREKLEE IDWVTFGVIL
301 KKVTPQSVNS GKTFSIWKLN DLRDLTQCVS LFLFGEVHKA LWKTEQGTVV
351 GILNANPMKP KDGSEEVCLS IDHPQKVLIM GEALDLGTCK AKKKNGEPT
401 QTVNLRDCEY CQYHVQAQYK KLSAKRADLQ STFSGGRIKP KFARRGTSLK
451 ERLCQDGFY GGVSSASYAA SIAAAVAPKK KIQTLSNLV VKGTNLIQIE
501 TRQKLGIPOK SLSCSEEFKE LMDLPTCGAR NLKQHLAKAS ASGIMGSPKP
551 AIKSIASAL LKQKQRMLE MRRRKSEIEQ KRFLQSSSEV ESPAVPSSSR
601 QPPAQPPRTG SEFPRLEGAP ATMTPKLGRG VLEGDDVLFY DESPPRPPLK
651 SALAEAKKLA AITKLRAKQ VLTKNPNNSI KKKQKDPQDI LEVKERVEKN
701 TMFSSQAED ELPARKKRRE QLAYLESEEF QKILKAKSKH TGILKEAEAE
751 MQERYFEPLV KKEQMEEKMR NIREVKCRV TCKTCAYTHF KLETCVSEQ
801 HEYHWHDGVK RFFKPCGNR SISLDRLPNK HCSNCGLYKW ERDGMKLVCH
851 LRTNF

```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h15, frame 2

TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347., N = 2, Score = 284, P = 7e-21

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7e-12

TREMBL:SCDNA52A_1 gene: "DNA52"; Saccharomyces cerevisiae DNA52 gene, complete cds., N = 2, Score = 201, P = 7.9e-12

TREMBLNEW:AC006234_6 gene: "F5H14.6"; Arabidopsis thaliana chromosome II BAC F5H14 genomic sequence, complete sequence., N = 2, Score = 211, P = 1.7e-15

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7.2e-12

>TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347.
Length = 593

HSPs:

Score = 284 (42.6 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 97/383 (25%), Positives = 186/383 (48%)

Query: 109 EKTNEELQEELRNLEQEMKALQEQLKVTIKQTASPARLQKSPEKSPRPPLKERRVQRIQ 168
E+ + +L+E + LQ Q+ +QE+ ++ + ++ AS + + PR P ++ RV +
Sbjct: 8 EENDLDLEE--KRLQRLNEIQEKKRLRSAQKEASSENAEVI--QVPRSPQVRLTVS 63

Query: 169 ESTCFSAE----LDVPALPRTKRVARTPKPSPDPKSSSRMTSAPSQP-----LQTIS 218
+ + + L + K V+ P P PK R+ A +Q L+T+
Sbjct: 64 SPSKLSKPKRLILGIDKGTGKDVSLGKGPGRGPLPKPFHERLAEARNQERKRSKDKLTKM 123

Query: 219 RNKPSGITRGQIVGTPGSSGETTQPI-C--VEAFSGLRLRRPRVSSSTEMNKKMTGRKLIR 275
+N+ R + + G S E P+ C ++ +S + +S + + G ++
Sbjct: 124 KNRKQSFQRKRNILEDGKSEEEKFPMKCDEIDPYSRQAIVIRYISDEVAKENIGGNQVYL 183

Query: 276 LSQIKEKMAREKLE--EID-WVTFGVILKKV-TPQSVNSGKTFSIWKLNDLRDLTQCVS 331
+ Q+ + + K E E+D +V G++ T ++VN K + + L DL+ +C
Sbjct: 184 IHQLLLVRAPKFEAPEVDNYVVMGIVASNSGTRETVNGNK-YCMLTLTDLKWQLEC--- 239

Query: 332 FLFGVHKALWKTEQGTVVGILNANPMKPKDGS-EEVCLSIDHPQKVLI-MGEALDLGTC 389
FLFG+ + WK + GTV+ +LN +KPK+ L +D VL+ +G + LG C
Sbjct: 240 FLFGKAFERYWKIQSGTVIALNPEVLKPKNPDIGRFSCLKLDSEYDVLLEIGRSKHLGYC 299

Query: 390 KAKKKNGEPTQTQTNLRDCEYCYHVQAQYKKLSAKRADLQSTFSGGRIPKKFARRGTS 449
+++K+GE C ++ R + C+YHV ++ + R + S+ + P+ ARR
Sbjct: 300 SSRKSGELCKHWLDKRGDVCEYHVDLAVQRSMSTRTEFASSMATMHEPR--ARR---- 353

Query: 450 KERLCQDGF--YGGVSSASYSIAAIVAPKKKIQT 484
++R GF Y+ G ++ ++A + +QT
Sbjct: 354 EKRFRGGGFQGYFAGEKYSAIPNAVAGLYDAEDAVQT 390

Score = 41 (6.2 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 12/43 (27%), Positives = 17/43 (39%)

Query: 453 LCQDGFYGGVSSASYSIAAIVAPKKKIQTTLNLVVKGTN 495
L +D S AS A++ K + SN + GTN
Sbjct: 465 LSKDSEIDSTKPSVLASFNASIMNPKSSLPSFSNSAILGTN 507

Score = 40 (6.0 bits), Expect = 8.9e-21, Sum P(2) = 8.9e-21
Identities = 13/26 (50%), Positives = 18/26 (69%)

Query: 536 LAKASASGIMGSPKPAIKSISASALL 561
LA +AS IM +PK ++ S S SA+L
Sbjct: 481 LASFNAS-IM-NPKSSLPSFSNSAIL 504

Pedant information for DKFZphtes3_2h15, frame 2

Report for DKFZphtes3_2h15.2

SEQ TCKTCAYTHFKLLETQVSEQHEYHWHGDKRFFKPCGNRSISLDRLPNKHCSNCGLYK


```
SEG .....
PRD eeecceeeeeeccccceeeccccccccceeeccccccccccccccccceec
COILS .....

SEQ ERDGMLKVCHLRTNF
SEG .....
PRD ccccccccccccccc
COILS .....
```

(No Prosite data available for DKFZphtes3_2h15.2)

(No Pfam data available for DKFZphtes3_2h15.2)

DKFZphtes3_2i5

group: testes derived

DKFZphtes3_2i5 encodes a novel 151 amino acid protein with weak similarity to. C.elegans cosmid F20D12.3

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans F20D12.3

many ATGs in front of the start of the ORF,
unspliced intron in 5' region?

Sequenced by EMBL

Locus: unknown

Insert length: 2142 bp

Poly A stretch at pos. 2121, polyadenylation signal at pos. 2102

```
1 GCAGTAAATA TGATATGAAA GAATTCTCTA ACTTGGGGGT GGCTTGTAAC
51 CTGTAATAAA AATATTGCTA AAATACCTTC TCTCACTTTG AAAAAGCATC
101 TGAGCAATCC TCAGTTATTG GTGAATTCTT ACCAGTGTTT AATTCCTCTC
151 TTTCGGTTAT GGTCTTAGTG TGGTTGTCCT GGTGTAGTAT TTCAAGAGGA
201 ACCTGCAGCA AGATGAAAAG AGAGTGGGAC TTGGAGCTAA GAACGTTTTT
251 GGCTTTAAGT GCTACGTTAA CTCATTAAAT TCTTAGTGAT CTTGGGGAAG
301 TCCCTCACC AGTGTGAGCC TCAGTTTCT TATCTAATAA GTAAGGATAA
351 TCTTACCAC CTTATTGCGG GGGCCCGAGG ATTACATGAT TGGTGTAAACA
401 GTAGCACCTT GTACATTGGA AAGGACTAAT ACCAGTGGAC TTAAACCTTG
451 GCTGGGCTTT GGAATTCTTG GTGGGACTTT TTAATCATGT AGATTCTCAG
501 GCCCTGCGCT GGCCTGTGGA ACCACAGACT CTATAGGTGG GCCCTTCCAG
551 AAGGCCTCAT GGGTGGTTCT CATGTGGAAC CTGTGTTGCA AGCCACTGCA
601 TGGTGTACT GCTATTAACA TTAAACTTA TATTTTCCTT ATGTGTGGA
651 TATACTGTG GTGTTTGGCC ATGTATACTT CATTTTACAT TTCTTAAAGA
701 ATAGAATGGA ATGGTTTAA GCACGCTACA TTGTCCAGGT TATACCCACA
751 GAAGAGCTGT TGTGTAACAG AATCAGCATC ATACCTGAAT CATTTGTACA
801 TTGCATATA GACTATGTCT AAGTAGAAGA TGCTATGAAA TCATGTCTGC
851 TGTGGGGCCA GGCATAATTA TGAATGTTAC TTAAGAGCAT AGGTGAGGTG
901 AGAAAAGGGA ATGTGACTAG TGTTTATGTA TTTTCTTGGT GTGGGATGAA
951 GTATAATTCT TTTTTTTTTT TCTCAACAAA GCAGTAAAC TAGAAAAGAAG
1001 GAGAACTCTT CCCTCAAGAA TGGCTGTACC TTCATATCTA GAGGCACATT
1051 AAAAAAAGA ACGTCTGTAC CTTAAAAATG GAGGTCATTT CATTGTGTTT
1101 ATTTTCAAGG TTGTTGTATG GCTCGGTCAG AACTTTCTGT TACCAGAAGA
1151 CACTCACATT CAGAATGCTC CATTTCAAGT GTGTTTACCA TCTTTACGGA
1201 ATGGCGGCCA CCTGCATATA AAAATAAAAC TTAGTGGAGA GATCACTATA
1251 AATACTGATG ATATTGATTT GGCTGGTGAT ATCATCCAGT CAATGGCATC
1301 ATTTTGTGCT ATTGAAGACC TTCAAGTAGA AGCGGATTTT CCTGTCTATT
1351 TTGAGGAATT ACGAAAGGTG CTAGTTAAGG TGGATGAATA TCATTCAAGT
1401 CATCAGAAGC TCAGTGCTGA TATGGCTGAT CATTCTAATT TGATCCGAAG
1451 TTTGCTGGTC GGAGCTGAGG ATGCTCGTCT GATGAGGGAC ATGAAAACAA
1501 TGAAGAGTCG TTATATGGAA CTCTATGACC TTAATAGAGA CTTGCTAAAT
1551 GGATATAAAA TTCGCTGTAA CAATCACACA GAGCTGTTGG GAAACCTCAA
1601 AGCAGTAAAT CAAGCAATTC AAAGAGCAGG TCGTCTGCGG GTTGGAAAAC
1651 CAAAGAACCA GGTGATCACT GCTTGTGCGG ATGCAATTCT AAGCAATAAC
1701 ATCAACACAC TGTTCAAAAT CATGCGAGTG GGGACAGCTT CTTCTAGGT
1751 GAGGAAAAAT CAGGTCATGA AGTTCCTGGC AAAGATTTTC TGTTAAAAAC
1801 CTATGCTGGT TTGCTTTGGA TCACACCCTG GTGAACCCCG GGTGCTAAGA
1851 ATGAAAAATA CCTTGGTGAG TTGTACAAAT TAAAGACAAA GAACTACATG
1901 TGAAGATAGA CTTGCTTTCT ATTTTAAAT CAGTAGTAGT ACTGTTGCTG
1951 AATAATACTA GGTTTTATG GAATAGGATG AATGCTTTTG AAGTATTAGG
2001 GCTTCAGAGT CCAATTTTGC TTATTTATGG TATATAAATA CATATTTTTT
2051 TCTTGAATTT GCAATTGAGT TTGTACTTTT CAAATAGATT ATCTACTTTT
2101 TCATTAAAT GTAAAGATGT TAAAAAATAA AAAAAAATAA AA
```

BLAST Results

No BLAST result

Medline entries

Peptide information for frame 3

```

1 MASFFAIEDL QVEADFVPVF EELRKVLVKV DEYHSVHVKL SADMDHNSL
51 IRSLLVGAED ARLMRDMKTM KSRYMELYDL NRDLLNGYKI RCNNHTELLG
101 NLKAVNQAIQ RAGRLRVGKP KNQVITACRD AIRSNNINTL FKIMRVGTAS
151 S

```

TREMBL:CEF20D12_1 gene: "F20D12.3"; *Caenorhabditis elegans* cosmid
F20D12., N = 1, Score = 173, P = 4.5e-12

Score = 173 (26.0 bits), Expect = 4.5e-12, P = 4.5e-12
Identities = 33/130 (25%), Positives = 72/130 (55%)

Query: 20 FEELRKVLVKVDEYHSVHQKLSADMDHNSLIRSLLVGAEDARLMRDMKMTKSRYMELYD 79
F+E ++L ++D V +L+A++ + ++ +++ AED+ + ++ + Y+ L
Sbjct: 569 FKEADEILEEIDPMTVEVRDLRLAELQERQAAVKEIIRAEDSAIDNIPDARKFYIRLKA 628

Query: 80 LNRDLLNGYKIRCNNHTELLGNLKAVNQAIQAGRLRVGKPKNQVITACRDAIRSNNINT 139
+ ++R NN + +L+ +N+ I+ RLRVG+P Q++ +CR AI +N
Sbjct: 629 NDAARQAAQLRWNNQERCVKSLRRLNKIENCRLRVGEPGRQIVVSCRSIAIDDNKQI 688

Query: 140 LFKIMRVGTA 149
+ KI++ G +
Sbjct: 689 ITKILQYGAS 698

Pedant information for DKFZphtes3_2i5, frame 3

Report for DKFZphtes3_2i5.3

```
[LENGTH]      151
[MW]           17304.07
[pI]           9.33
[HOMOL]        TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12. 2e-12
[KW]           Alpha_Beta
```

SEQ MASFFAIEDLQVEADFVPVYFEELRKVLVKVOEYHSVHQKLSADMAHDSNLRSLLVGAED
PRD cceeeehhhhhhccccchhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ ARLMRDMKTMKSRYMELYDLNRDLLNGYKIRCNNHTELLGNLKAVNQAIQRAGRLRVGKP
PRD hhhhcccchhhhhhheeeccchhhhhhheeeeecchhhhhhhhhhhhhhhhhhhcccccc

SEQ KNQVITACRDIAIRSNNTLFKIMRVGTASS
PRD cceeeehhhhhhccccceeecccccccccc

(No Prosite data available for DKFZphtes3_2i5.3)

(No Pfam data available for DKFZphtes3_2i5.3)

DKFZphtes3_2119

group: testes derived

DKFZphtes3_2119 encodes a novel 166 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, no EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1079 bp

Poly A stretch at pos. 1053, polyadenylation signal at pos. 1038

```
1 CCACAGGACA CACTGTTCCT AGGGCACAGA CACCCTGGGC TTTGGTTGGG
51 TCTTGGCCTC CAGGTAGGGC CCTGTTGGGC AGCGGGCAGC AACTCCTGAG
101 ACACACTACTGT GATTCTTGGT GGTGGCTGTG GTAAAAAACC TGCAGGGCTA
151 GAGTTTGGGG TGAGATTCAG CAGTAACTGT GGCCTCTCCT AGTGACAGTA
201 TGTCACCTCCC ACTCCCAGCA CGCATGCCCA CAGGCCACGG CCTCCACATC
251 ACAAACCCCC CACCAAGTTG CCCATCTATG GAGCAGCTCC CATACGGCAG
301 GGTCAGGCTC TTACCTCCAC CTCCAGGGCA CAGACAGGGG GAGCTCTGTC
351 TCACTGTAAG GCAATGAGGA GAGTTGAGGG CCCAGACCAG GCTAGGGGCC
401 ATCCCCTTTC CCGAGCAGGC CTCAGGGAAG GACCAGCCCC ATTCCCCTCT
451 GACCTAGGTC TTAGCCCAGG AGCCTGCATA GGAAGAAAAG GACAGACAGG
501 GCCTCCTTAC TGGCTGACAC TCAGGAGGGG CTGGGGCAAG AGAGCAGAGG
551 GAGCGCAGGG CCAGGCAGGG GCTGCTGAGG ATCCATGGGA GCTCAGGGTG
601 CACAAGGGGG CTGCCCTTCC TGGGCTGCAG GCAGCATCCC TATGGGAGCT
651 GAGAAAGTCC AATCCTGAGA TGGGACAGTG CTGCCCAGGG GTGTGTGGCT
701 GGGCCCTGAC AACAGTCTCC CAAAAGTGA CCACATCACC AGGCTCAGTT
751 CCAGGAAGGC TGAGAAGTGC CCAGTACACT GAGGATGCAC CTCAGTTACA
801 TAAAAATAAT GAAACTGGAG TACTAACGTA CAGTTTAAAG GTTATAGTTA
851 CTATTTTAT ATGATATACT AGTAATTTT GAATAGGGTA AACTTTAGGT
901 GTTTTGACAC CAAAAGAAAA CTACATGAGT TCATGCATGT GTTAAATTGC
951 TTTACTGTAG TAATCATTTA CATGTATATG TATATATGAA TATAATTATG
1001 GGCTCATTAA ATTTAAATAT TATAAATAGG TGACAAAGAA TAAAGTTAAC
1051 TGGAAAAAAA AAAAAAATAA AAAAAAATAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 364 bp to 861 bp; peptide length: 166
Category: putative protein
Classification: no clue

```
1 MRRVEGPDQA RGHPLSRAGL REGPAPFSPD LGLSPGACIG KKGQTGPPYW
51 LTLRRGWGKR AEGAQQQAGA AEDPWELRVH KGAALPGLQA ASLWELRKS
101 PEMQCCPGV CGWALTTVSP KVTTSPPGSPV GRLRSAQYTE DAPQLHKINE
151 TGVLTYSKLV IVTIFI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2119, frame 1

Report for DKFZphtes3_2119.1

{LENGTH} 166
{MW} 17691.35
{pI} 9.54
{KW} All_Beta
{KW} LOW_COMPLEXITY 7.23 %

SEQ MRRVEGPDQARGHPLSRAGLREGPAPFPSPDLGLSPGACIGKKGQTGPPYWLTLRRGWGKR
SEG
PRD ccc

SEQ AEGAQQAGAAEDPWELRVHKGAAALPGLQAASLWELRKSNPEMGQCCPGVCGWALTTVSP
SEG xxx
PRD ccc

SEQ KVTTSFGSVPGRLRSAQYTEDAPQLHKINETGVLTYSLKVIVTIFI
SEG
PRD ccc

(No Prosite data available for DKFZphtes3_2119.1)

(No Pfam data available for DKFZphtes3_2119.1)

DKFZphtes3_2ml8

group: nucleic acid management

DKFZphtes3_2ml8 encodes a novel amino acid protein, with similarity to mouse Dhml.

The protein seems to play a role in nucleotide metabolism, RNA metabolism, but also in DNA repair and cell cycle. The yeast homologue is a DNA strand exchange protein required for sporulation and homologous recombination.

The novel protein can find application as multifunctional nuclease / exoribonuclease.

nearly identical to mouse Dhml

complete cDNA, complete cds, start at Bp 42, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3022 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2981

```
1 CTCGTCAGCC GGTCCGCCGC CGCCTCCAGC CGTGTGCCGC TATGGGAGTC
51 CCGGCGTTCT TCCGCTGGCT CAGCCGCAAG TACCCGTCCA TCATAGTCAA
101 CTGCGTGGAA GAGAAGCCAA AAGAATGCAA TGGTGTAAG ATTCCAGTTG
151 ATGCCAGTAA ACCTAATCCA AATGATGTGG AGTTTGATAA TCTGTATTTG
201 GATATGAATG GAATCATCCA TCCCTGTACT CATCCTGAAG ACAAACCAGC
251 ACCAAAAAAT GAAGATGAAA TGATGGTTGC AATTTTGTAG TACATTGACA
301 GACTTTTCAG TATTGTAAGA CCAAGAAGAC TTCTCTACAT GGCAATAGAT
351 GGAGTGGCAC CACGTGCTAA AATGAACCAG CAGCGTTCAG GGAGGTTTCAG
401 GGCATCAAAA GAAGGAATGG AAGCAGCAGT CGAGAAGCAG CGAGTCAGGG
451 AAGAAATATT GGCAAAAGGT GGCTTTCTTC CTCCAGAAGA AATAAAAGAA
501 AGATTTGACA GCAACTGTAT TACACCAGGA ACTGAATTCA TGGACAATCT
551 TGCTAAATGC CTTGCTATT ACATAGCTGA TCGTTTAAAT AATGACCCTG
601 GGTGGAAGAA TTTGACAGT ATTTTATCTG ATGCTAGTGC TCCTGGTGAA
651 GGAGAACATA AAATCATGGA TTACATTAGA AGGCAAGAG CCCAGCCTAA
701 CCATGACCCA AATACTCATC ATTGTTTATG TGGAGCAGAT GCTGATCTCA
751 TTATGCTTGG CTTGCCCACA CATGAACCGA ACTTTACCAT TATTAGAGAA
801 GAATCAAAAC CAAACAAGCC CAAACCATGT GGTCTTGTGA ATCAGTTTGG
851 ACATGAGGTC AAAGATTGTG AAGGTTTGCC AAGAGAAAAG AAGGGAAAGC
901 ATGATGAAC TCCGATAGT CTTCTTGTG CAGAAGGAGA GTTTATCTTC
951 CTTGCGCTTA ATGTTCTTCG TGAGTATTTG GAAAGAGAAC TCACAATGGC
1001 CAGCCTACCA TTCACATTG ATGTTGAGAG GAGCATTGAT GACTGGGTTT
1051 TCATGTGCTT CTTTGTGGGA AATGACTTCC TCCCTCATTT GCCATCGTTA
1101 GAGATTAGGG AAAATGCAAT TGACCGTTTG GTTAACATAT ACAAATATGT
1151 GGTACACAAA ACTGGGGGTT ACCTTACAGA AAGTGGTTAT GTCAATCTGC
1201 AAAGAGTACA GATGATCATG TTAGCAGTTG GTGAAGTTGA GGATAGCATT
1251 TTTAAAAAGA GAAAGGATGA TGAGGACAGT TTTAGAAGAC GACAGAAAGA
1301 AAAAAGAAAG AGAATGAAGA GAGATCAACC AGCTTTCACT CCTAGTGGAA
1351 TATTAATCC TCATGCCTTG GGTCAAGAA ATTCAACAGG TTCTCAAGTA
1401 GCCAGTAATC CGAGACAAGC AGCCTATGAA ATGAGGATGC AGAATAACTC
1451 TAGTCCTTCG ATATCTCCTA ATACGAGTTT CACATCTGAT GGCTCCCCGT
1501 CTCCATTAGG AGGAATTAAG CGAAAAGCAG AAGACAGTGA CAGTGAACCT
1551 GAGCCAGAGG ATAATGTCAG CTTATGGGAA GCTGGCTGGA AGCAGCGGTA
1601 CTACAAGAAC AAATTGATG TGGATGCAGC TGATGAGAAA TTCCGTCGGA
1651 AAGTTGTGCA GTCGTACGTT GAAGGACTTT GCTGGGTTCT TAGATATTAT
1701 TACCAGGGCT GTGCTTCCTG GAAGTGGTAT TATCCATTTC ATTATGCACC
1751 ATTTGCTTCA GACTTTGAAG GCATTGCAGA CATGCCATCT GATTTTGAGA
1801 AGGGTACGAA ACCGTTTAAA CCACTAGAAC AACTTATGGG GGTATTTCCA
1851 GCTGCAAGTG GTAATTTTCT ACCTCCATCA TGGCGGAAGC TCATGAGTGA
1901 TCCTGATTCT AGTATAATTG ACTTCTATCC TGAAGATTTT GCTATTGATT
1951 TGAATGGGAA GAAATATGCA TGGCAAGGTG TTGCTCTCTT GCCATTCTGT
2001 GATGAGCGAA GGCTACGAGC TGCCCTAGAA GAGGTATACC CAGACCTCAC
2051 TCCAGAAGAG ACCAGAAGAA ACAGCCTTGG AGGTGATGTC TTATTTGTGG
2101 GGAAACATCA CCCACTCCAT GACTTCATTT TAGAGCTGTA CCAGACAGGT
2151 TCCACAGAGC CAGTGGAGGT ACCCCCTGAA CTATGTCATG GGATTCAAGG
2201 AAAGTTTTCT TTGGATGAAG AAGCCATTCT TCCAGATCAA ATAGTATGTT
2251 CTCTGTCTCC TATGTTAAGG GATCTGACAC AGAACACTGT AGTCAGTATT
2301 AATTTTAAAG ACCCACAGTT TGCTGAAGAT TACATTTTAA AAGCTGTAAT
2351 GCTTCCAGGA GCAAGAAAGC CAGCAGCAGT ACTGAAACCT AGTGACTGGG
2401 AAAAATCCAG CAATGGACGG CAGTGGAAAG CTCAGCTTGG CTTTAACCGT
2451 GACCGGAGGC CTGTGCACCT GGATCAGGCA GCCTTCAGGA CTTTGGGCCA
2501 TGTGATGCCA AGAGGCTCAG GAACTGGCAT TTACAGCAAT GCTGCACCAC
2551 CACCTGTGAC TTACCAGGGA AACTTATACA GGCCGCTTTT GAGAGGACAA
2601 GCCCAGATTC CAAACTTAT GTCAATATAG AGGCCCCAGG ATTCCTGGCG
2651 AGGTCTCTCT CCCCTTTTCC AGCAGCAAGG GTTTGACAGA GGCGTTGGGG
```



```

2701 CTGAACCTCT GCTCCCATGG AACCGGATGC TGCAAACCCA GAATGCAGCC
2751 TTCCAGCCAA ACCAGTACCA GATGCTAGCT GGGCCTGGTG GGTATCCACC
2801 CAGACGAGAT GATCGTGGAG GGAGACAGGG ATATCCCAGA GAAGGAAGGA
2851 AATACCTTTT GCCACCACCC TCAGGAAGAT ACAATTGGAA TTAAGCTTTT
2901 GTAAAGCTTT CCCAAATCCT TTCATCATTC TACAGTTTGA TGCTATTTGT
2951 GGAAAGATTT CTTCTCAAG TAGTAGTTT TAATAAACT ACAGTACTTT
3001 GTGTAAGAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

95192042:

Characterization of cDNA encoding mouse homolog of fission yeast dhpl+
gene: structural
and functional conservation.

97361754:

Cloning and characterization of mouse Dhml2 cDNA, a functional homolog
of budding yeast
SEPI.

Peptide information for frame 3

ORF from 42 bp to 2891 bp: peptide length: 950
Category: strong similarity to known protein

```

1  MGVPAFFRWL SRKYPSIIVN CVEEKPECN GVKIPVDASK PNPNDVEFDN
51  LYLDNMGIH PCTHPEDKPA PKNEDEMMVA IFEYIDRLFS IVRPRRLLYM
101 AIDGVAPRAK MNQQRSSRRF ASKEGMEAAV EKQVRVEIL AKGGFLPPEE
151 IKERFDSNCI TPGTEFMDNL AKCLRYIAD RLNNDPGWKN LTVILSDASA
201 PGEGERKIMD YIRRQRAQPN HDPNTHHCLC GADADLIMLG LATHEPNFTI
251 IREEFKPNKP KPCGLCNQFG HEVKDCEGLP REKKGKHDEL ADSLPCAEGE
301 FIFLRLNVLR EYLERELTMA SLPFTFDVER SIDDWVFMCF FVGNDFLPHL
351 PSLEIRENAI DRLVNIYKNV VHKTGGYLTE SGYVNLQVRQ MIMLAVGEVE
401 DSIFKKRKDD EDSFRRRQKE KRKRMRDQD AFTPSGILTP HALGSRNSPG
451 SQVASNPQQA AYEMRMQNN SPSISPNTSF TSDGSPSPLG GIKRKAEDSD
501 SEPEPEDNVR LWEAGWKQRY YKNKFDVDA DEKFRKRVVQ SYVEGLCWVL
551 RYYYQGCA SWYYPFHYAP FASDFEGIAD MPSDFEKGTK PFKPLEQLMG
601 VFPAASGNFL PPSWRKLMSD PDSSIIDFYP EDFAIDLNGK KYAWQGVALL
651 PFVDERRLRA ALEEVYPDLT PEETRRNSLG GDVLFVGKHH PLHDFILELY
701 QTGSTEPVEV PPELCHGIQG KFSLDDEAIL PDQIVCSFVP MLRDLTQNTV
751 VSINFKDPQF AEDYIFKAVM LPGARKPAAV LKPSDWEKSS NGRQWKPQLG
801 FNRDRRPVHL DQAAFRTLGH VMPRGSGTGI YSNAAPPVPT YQGNLYRPLL
851 RGQAQIPKLM SNMRPQDSWR GPPPLFQQQR FDRGVGAEP LPPWNRMLQTQ
901 NAAFQPNQYQ MLAGPGGYPP RRDDRGRGQG YPREGRKYPL PPPSGRYNWN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2ml8, frame 3

PIR:I49635 mouse Dhml protein - mouse, N = 1, Score = 4765, P = 0

PIR:S43891 dhpl protein - fission yeast (Schizosaccharomyces pombe), N = 3, Score = 1172, P = 2e-197

PIR:S20126 exoribonuclease RAT1 (EC 3.1.11.-) - yeast (Saccharomyces cerevisiae), N = 2, Score = 1146, P = 3.8e-175

PIR:S72531 exonuclease II - fission yeast (Schizosaccharomyces pombe), N = 4, Score = 622, P = 4.2e-125

>PIR:I49635 mouse Dhml protein - mouse
Length = 947

HSPs:

Score = 4765 (714.9 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 884/930 (95%), Positives = 895/930 (96%)

```

Query:   1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDFEDNLYLDMNGIIH 60
          MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDFEDNLYLDMNGIIH
Sbjct:   1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDFEDNLYLDMNGIIH 60

Query:  61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSRFR 120
          PCTHPEDKPAPKNEDEMMVAIFEYIDRLF+IVRPRRLLYMAIDGVAPRAKMNQQRSRFR
Sbjct:  61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFNIVRPRRLLYMAIDGVAPRAKMNQQRSRFR 120

Query:  121 ASKEGMEA AVEKQVRREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD 180
          A K GMEA AVEKQVRREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD
Sbjct:  121 AIKGGMEA AVEKQVRREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD 180

Query:  181 RLNDPBGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG 240
          RLNDPBGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG
Sbjct:  181 RLNDPBGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNQDPNTHHCLCGADADLIMLG 240

Query:  241 LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300
          LATHEPNFTIIREEFKPNKPKPC LCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
Sbjct:  241 LATHEPNFTIIREEFKPNKPKPCALCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300

Query:  301 FIFLRNLVREYLERELTMASLPFTFDVERSIDDWVFMCFVGNDFLPHLPSLEIRENAI 360
          FIFLRNLVREYLERELTMASLPF FDVERS DDW FMCFFVGNDFLPHLPSLEIRE AI
Sbjct:  301 FIFLRNLVREYLERELTMASLPFFFDVERSNDWVFMCFVGNDFLPHLPSLEIREGAI 360

Query:  361 DRLVNIYKNVVHKTGGYLTESGYVNLQVQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE 420
          DRLVNIYKNVVHKTGGYLTESGYVNLQVQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE
Sbjct:  361 DRLVNIYKNVVHKTGGYLTESGYVNLQVQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE 420

Query:  421 KKKRMKRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPNTSF 480
          KKKRMKRDQPAFTPSGILTPHALGSRNSPG QVASNPRQAAYEMRMQ NSSPSISPNTSF
Sbjct:  421 KKKRMKRDQPAFTPSGILTPHALGSRNSPGCQVASNPRQAAYEMRMQNNSSPSISPNTSF 480

Query:  481 TSDGSPSPGLGKIRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRKRVVQ 540
          SDGSPSPGLGI+RKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRKRVVQ
Sbjct:  481 ASDGSPSPGLGIRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRKRVVQ 540

Query:  541 SYVEGLCWVLRYYYQGCASWKWYYPFHYPFASDFEGIADMPDSEKGTKPKFPLEQLMG 600
          SYVEGLCWVLRYYYQGCASWKW YPFHYAPFASDFEGIADM S+FEKGTKPKFPLEQLMG
Sbjct:  541 SYVEGLCWVLRYYYQGCASWKWLYPFHYAPFASDFEGIADMSSEFEKGTKPKFPLEQLMG 600

Query:  601 VFPAASGNFLPSPWRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660
          VFPAASGNFLPP+WRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
Sbjct:  601 VFPAASGNFLPPTWRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660

Query:  661 ALEEVPDLTPEETRNSLGGDVLFGVGHPLHDFILELYQTGSTPEVPELCHGIQG 720
          ALEEVPDLTPEE RRNSLGGDVLFGVGHPLHDFILELYQTGSTPEV+VPELCHGIQG
Sbjct:  661 ALEEVPDLTPEENRRNSLGGDVLFGVGHPLHDFILELYQTGSTPEVDVPELCHGIQG 720

Query:  721 KFSLDEEAILPDQIVCSVPVMLRDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAV 780
          FSLDEEAILPDQ VCSVPVMLRDLTQNT VSINFKDPQFAEDY+FKA MLPARKPA V
Sbjct:  721 TFSLDEEAILPDQTVCSVPVMLRDLTQNTAVSINFKDPQFAEDYVFKAAMLPGARKPATV 780

Query:  781 LKPSDWEKSSNGRWKPOLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPVPT 840
          LKP DWEKSSNGRWKPOLGFNRDRRPVHLDQAAFRTLGHV PRGSGT +Y+N A P
Sbjct:  781 LKPGDWEKSSNGRWKPOLGFNRDRRPVHLDQAAFRTLGHVTPRGSGTSVYNTALLPAN 840

Query:  841 YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEPPLPWNRMLQTQ 900
          YQGN YRPLLRGQAQIPKLMSNM RP+DSWRGPPPLFQQ RF+R VGAEPPLPWNRM+Q Q
Sbjct:  841 YQGNLYRPLLRGQAQIPKLMSNM RPKDSWRGPPPLFQQHREFERSVGAEPPLPWNRMIQNQ 900

Query:  901 NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ 929
          NAAFQPNQYQML GPGGYPPRRDD RGGRQ
Sbjct:  901 NAAFQPNQYQMLGGPGGYPPRRDDHRGGRQ 930

```

Pedant information for DKFZphtes3_2ml8, frame 3

Report for DKFZphtes3_2ml8.3

[LENGTH]	950
[MW]	108582.68
[pI]	7.26
[HOMOL]	PIR:I49635 mouse Dhml protein - mouse 0.0
[FUNCAT]	08.01 nuclear transport [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]	04.01.04 rrna processing [S. cerevisiae, YOR048c] 1e-123


```

[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR048c] 1e-123
[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YGL173c] 3e-79
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGL173c] 3e-79
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL173c] 3e-79
[PIRKW] nucleus 1e-126
[PIRKW] hydrolase 1e-122
[PIRKW] exoribonuclease 1e-122
[PROSITE] MYRISTYL 7
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 4
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 6.21 %

```

```

SEQ MGVPAPFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDFEONLYLDMNGIIH
SEG .....
PRD cccchhhhhhhhhccceeeeeccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVPRRLLYMAIDGVAPRAKMNQQRSRFR
SEG .....
PRD cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhccceeeeeccccchhhhhhhhhhhhh
MEM .....

SEQ ASKEGMAAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhh
MEM .....

SEQ RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHCLCGADADLIMLG
SEG .....
PRD hccccccccceeeeeccccccccchhhhhhhhhhhhhhhccccccccccccccccccccceec
MEM .....

SEQ LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
SEG .....
PRD cccccccccccccccccccccccccceccccccccccccccccchhhhhhhhhcccccccccc
MEM .....

SEQ FIFLRLNVLEYLELTMASLPFTFDVERSIDDWVFCFFVGNDFLPHLPSLEIRENAI
SEG .....
PRD cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ DRLVNIYKNVVKHTGGYLTESGYVNLQRVQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE
SEG .....
PRD hhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ KRKRMRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAAYEMRMQNNSSPSISPNTSF
SEG .....
PRD hhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ TSDGSPSPGLGKIRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADERFRKRVVQ
SEG .....
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ SYVEGLCWVLRYYQGCASWKWYYPFHYAPFASDFEGIADMPDSEKGTKPKPLEQLMG
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
SEG .....
PRD hccccccccccccccccccccccccceccccccccccccccccccccccccccccccccchhhhh
MEM .....

SEQ ALEEVYPLDTPETRRNSLGGDVLFGKHHPLHDFILELYQTGSTPEVPELCHGIQ
SEG .....
PRD hhhhccccchhhhhhhccccccccceccccchhhhhhhhhhhhhcccccccccccccccccccc
MEM .....

SEQ KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV
SEG .....

```



```

PRD      cccccceeeccccceccccccccccccccccccccccccchhhheccccccccccce
MEM      .....

SEQ      LKPSDWEKSSNGRQWKPLGFRDRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPFPVT
SEG      .....
PRD      eccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccc
MEM      .....

SEQ      YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEP LLPWNRMLQTQ
SEG      .....
PRD      cccccchhhhhccccchhhhhccccccccccccccccchhhhhccccccccccccchhhhh
MEM      .....

SEQ      NAAFQPNQYQMLAGPGGYPPRRDDRGRQGYPREGRKYPLPPPSGRYNWN
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      hcccccccccecccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphtes3_2ml8.3

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	247->251	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	477->481	ASN_GLYCOSYLATION	PDOC00001
PS00002	826->830	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	675->679	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00005	559->562	PKC_PHOSPHO_SITE	PDOC00005
PS00005	613->616	PKC_PHOSPHO_SITE	PDOC00005
PS00005	674->677	PKC_PHOSPHO_SITE	PDOC00005
PS00005	868->871	PKC_PHOSPHO_SITE	PDOC00005
PS00005	944->947	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	501->505	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	619->623	CK2_PHOSPHO_SITE	PDOC00006
PS00006	624->628	CK2_PHOSPHO_SITE	PDOC00006
PS00006	670->674	CK2_PHOSPHO_SITE	PDOC00006
PS00006	723->727	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->667	TYR_PHOSPHO_SITE	PDOC00007
PS00008	125->131	MYRISTYL	PDOC00008
PS00008	375->381	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00008	825->831	MYRISTYL	PDOC00008
PS00008	829->835	MYRISTYL	PDOC00008
PS00008	926->932	MYRISTYL	PDOC00008
PS00009	638->642	AMIDATION	PDOC00009
PS00009	934->938	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2ml8.3)

DKFZphtes3_2m20

group: testes derived

DKFZphtes3_2m20 encodes a novel 183 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

group: unknown

DKFZphtes3_2m20 encodes a novel

amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

EST hits are only from testis or uterus librarys
remaining intron in 3' UTR see EST-BLAST

Sequenced by EMBL

Locus: unknown

Insert length: 1341 bp

Poly A stretch at pos. 1320, polyadenylation signal at pos. 1300

```
1  GCAATCCAGG AGCTGAATGG TAACTCTTCC ACAAGCGAAA ACTGTTTCGTG
51 AATACAAGCA AAAGGCCCCC CAAGAGGACC CCTGATATGA TCCAGCAGCC
101 TCGGGCCCCG CTGGTGTGGG AGAAGGCTTC TGGTGAAGGA TTTGGCAAAA
151 CCGCCCGCTAT TATACAGCTC GCTCCTAAAG CTCCTGTGGA CCTGTGTGAG
201 ACAGAGAAAC TGAGGGCAGC CTTCTTTGCA GTCCCGTTGG AAATGAGAGG
251 GTCCTTCCTG GTGCTGCTCC TGAGGGAATG CTTCCGAGAC CTGAGCTGGC
301 TGGCACTCAT CCATAGCGTC CGTGGGGAGG CGGGGCTGCT GGTGACGAGT
351 ATCGTCCCGA AGACCCCGTT TTTCTGGGCC ATGCACATCA CTGAGGCTCT
401 GCACCCAGAA ATGCAGGCTC TGTTTAGCAC CCTGGCTCAG GCGGAGGAGC
451 AGCAGCCCTA CCTGGAGGCT CCACCGTTAT GCGCGGGACT CGCTGTCTGG
501 CAGAGTACCA CCTGGGGGAT TATGGACACG CCTGGAACAG GTGTTGGGTG
551 CTGGACAGGG TGGACACCTG GGCTGTGGTC ATGTTTCATTG ATTTTGGACA
601 GTTGGCCACC ATCCCTGTGC AGTCTCTGCG CCAGCTAGAC AGCGACGACT
651 TCTGGACCAT CCCACCCCTG ACTCAGCCAT TCATGCTGGA GAAAGACATT
701 TTGAGTTCGT ATGAGGTTGT CCATCGAATC CTCAAAGGGA AAATCACTGG
751 TGCTTTGAAC TCGGCGGTAA CTGCTCCTGC ATCTAACTTG GCTGTTGTCC
801 CTCCACTCCT GCCCTTGGGG TGTCTGCAGC AGGCTGCTGC CTAGGCCTGG
851 ACACATTGCA CATCCTAAAG TTTGAAGAGT CTAATAACG GGGCTTCCCT
901 CAGCATGTTC CCTCTCCTGT TTGCCACGGA TCCAGAGCCA CCTGCCCTGT
951 CTTCTCGTAC CCCTTTCACT CTTGAGGCCT GGGAGGTGAA AAAGGCCAGA
1001 CTGTGCCCAG GATTGATTCA ATTTTGCTTT TACTCCCAGC TTCCCTCTCA
1051 AAAGAGAGTG AAGTCTCATT TGTCATGTGT CTTCACTTCC CCAACTTGCC
1101 ATGAACATTT GAACCAAACA TAGGAAACTA CCATTAGGTT GAAAGCCTGA
1151 GGCAGCTGGG ATGGTCTTTC TTGTGTCTCT TCTTGCACC CCAGAGCATG
1201 ATATAAGTGG TCCTAACAGA TTCTGGATAA TGGAGAAGCC CTCTGCTGGT
1251 TTTCCCTGGCA TTCCATGTAG AATAGGTAGA GAATATTTAA CCAATGAGCA
1301 AATAAATGTT GGCATGTTTC ATGAAAAAAA AAAAAAAA A
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 479 bp to 841 bp; peptide length: 121
 Category: questionable ORF
 Classification: no clue

1 MRGTRCLA EY HLG DYGHAWN RCWVLD RVD T WAVVMFIDFG QLATIPVQSL
 51 RQLDSDDFWT IPPLTQPFML EKDILSSYEV VHRILKGKIT GALNSAVTAP
 101 ASNLAVVPPL LPLGCLQAAA A

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 87 bp to 635 bp; peptide length: 183
 Category: putative protein
 Classification: no clue

1 MIQOPRAPLV LEKASGEGFG KTAATLQ LAP KAPVDLCETE KLRAAFFAVP
 51 LEMRGSFLVL LLRECFRDL S WLALIH SVRG EAGLLVTSIV PKTPFFWAMH
 101 ITEALHQNMQ ALFSTLAQAE EQQPYLEAPP LCAGLAVWQS TTWGIMDTGP
 151 TGVGCWTGWT PGLWSCSLIL DSWPPSLCSL CAS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2m20, frame 2

Report for DKFZphtes3_2m20.2

[LENGTH] 121
 [MW] 13436.69
 [pI] 5.81
 [KW] Alpha_Beta

SEQ MRGTRCLA EYHLG DYGHAWNRCWVLD RVD T WAVVMFIDFG QLATIPVQSLRQLDSDDFWT
 PRD cccchhhhhcc

SEQ IPPLTQPFMLEKDILSSYEVVHRILKGKITGALNSAVTAPASNLAVVPPLPLGCLQAAA
 PRD cccccchhhhhhhcchhhhhhhhhccccccccchhhhhcccccccccccccccccccccccc

SEQ A
 PRD c

(No Prosite data available for DKFZphtes3_2m20.2)

(No Pfam data available for DKFZphtes3_2m20.2)

Pedant information for DKFZphtes3_2m20, frame 3

Report for DKFZphtes3_2m20.3

[LENGTH] 183
 [MW] 19971.49
 [pI] 5.31
 [KW] Alpha_Beta


```
SEQ  MIOQPRAPLVLEKASGEGFGKTAAIQLAPKAPVDLCETEKLRAAFFAVPLEMRGSFLVL
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  LLRECFRDLISWLALIHVSVRGEAGLLVTSIVPKTPFFWAMHITEALHQNMQALFSTLAQAE
PRD  hhhhhhhcchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc
SEQ  EQQPYLEAPPLCAGLAVWQSTTWGIMDTPGTGVGCWTGWTPGLWSCSLILDSWPPSLCSL
PRD  hhhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  CAS
PRD  ccc
```

(No Prosite data available for DKFZphtes3_2m20.3)

(No Pfam data available for DKFZphtes3_2m20.3)

DKFZphtes3_2n9

group: testes derived

DKFZphtes3_2n9 encodes a novel 184 amino acid protein with very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

on genomic level encoded by HS1186N24, no splice pattern but EST matches

Sequenced by EMBL

Locus: unknown

Insert length: 1000 bp

Poly A stretch at pos. 988, polyadenylation signal at pos. 970

```

1 CAACTTTTAA AAGATGTGAA TTGGACAGCC AGACTTGCTT ATTTGTCTGA
51 TATCTTCAGT ATTTTAAAT GATCTTAATG CTTCTATGCA AGGGAAGAAT
101 GCAACTTATT TTCAATGGC AGATAAAGTT GAAGGACAAA AACAGAAGTT
151 AGAAGCTTGG AAAACAGAA TTTCTACAGA TTGTATGAC ATGTTTCATA
201 ATTTAACAAC AATTATCAAT GAAGTAGGTA ATGATCTTGA TATTGCACAT
251 CTGCGAAAAA TTATCAGTGA ACATCTTACA AATTGTGTAG AATGTTTGA
301 ATTTTATTTT CCATCAAAG AAGATCCACG CATAGGAAAT TTGTGGATCC
351 AAAATCCATT TCTTTCATCA AAAGATAACT TAAATTTAAC TGTAACTCTA
401 CAGGATAAGT TGTGAAGCT GGCTACCGAC GAAGGATTGA AAATCAGTTT
451 TGAAAAATACA GCATCACTTC CTTCAATTTG GATAAAAGCT AAAAATGACT
501 ATCCTGAGCT TGCTGAGATT GCTTTAAAT TGCTGCTTCT TTTCCCTCA
551 ACATACCTCT GTGAGACCGG ATTCTCTACT TTAAGTGTTA TTAACACAAA
601 ACATAGAAAC AGTTTAAATA TACATTATCC CCTGAGGTAG CATTGTCATC
651 AATCCAACTT AGATTAGACA AATTAACAAG CAAGAAGCAA GCTCACTTAT
701 CACATTAAAA GCTTTAAATA TTGATATGTA AGGTATTGGT TCAAAGTATG
751 CATATAAGCA TTGAGTGTGA GGAATTTGCT ATTTCACTTT AAACTTTCTG
801 TCTAGTTTACA GTTATGGAAG TATGAGAAGT TATGAGTGAA ACAGCAATTT
851 TCTATATAAA TTGCCTATAT GTATATTTTC AATTAAGAAAT GTGTACAGTT
901 TTTATAATTC TATTTTCTCT CATATTTGTC GTATTTATTA AAATATAATT
951 TTAATCTGTG TGATTCTAAT ATTAAACAT TTGATCTTAA AAAAAAAAAA

```

BLAST Results

Entry HS1186N24 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 1186N24
 Score = 4921, P = 5.8e-215, identities = 989/992

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 86 bp to 637 bp; peptide length: 184
 Category: similarity to unknown protein
 Classification: no clue

```

1 MQGKNATYFS MADKVEGQKQ KLEAWKNRIS TDCYDMFHNL TTIINEVGND
51 LDIAHLRRVI SEHLTNLLEC FEFYFPSKED PRIGNLWION PFLSSKDNLM
101 LTVTLQDKLL KLATDEGLKI SFENTASLPS FWIKAKNDYP ELAEIALKLL
151 LLFPSTYLCE TGFSTLSVIK TKHRNSLNIH YPLR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2n9, frame 2

TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence., N = 1, Score = 94, P = 0.042

>TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence.
Length = 533

HSPs:

Score = 94 (14.1 bits), Expect = 4.3e-02, P = 4.2e-02
Identities = 39/177 (22%), Positives = 75/177 (42%)

```
Query:      1 MOGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLD-IAHLRKV 59
             +QG +      M D +      KL W+ ++ +      F L      + L+ I + ++
Sbjct:     354 LQGHSQLVTQMYDLIRAFKAKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYIPKIAEL 413

Query:      60 ISEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLK 119
             +E      L + F+ Y  + + +      +PF + D+++      LQ +++ L + LK
Sbjct:     414 KTEFQKRLSD-FKLY---ESELTL----FSSPFSTKIDSVH--EELQMEVIDLQCNTVLK 463

Query:      120 ISFENTASLPSEFWIKAKNDYPXXXXXXXXXXXXFPSTYLCETGFSTLSVIKTKHRNSL 177
             ++      +P F+      YP      F STY+CE FS + + KTK+ + L
Sbjct:     464 TKYDKVG-IPFQKYLWGSYPKYKHHKAKILSMFGSTYICEQLFSIMKLSKTKYCSQL 520
```

Pedant information for DKFZphtes3_2n9, frame 2

Report for DKFZphtes3_2n9.2

```
[LENGTH]      184
[MW]           21203.53
[pI]           6.52
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      6.52 %
```

```
SEQ  MOGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLDIAHLRKVI
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhcchhhhhccceeeccccccccchhhhhhhhh

SEQ  SEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLKI
SEG  .....
PRD  hhhhhhhhhhhhhcccccceeeccccccccceeeehhhhhhhhhhhcccee

SEQ  SFENTASLPSEFWIKAKNDYPELAELKLLLPSTYLCETGFSTLSVIKTKHRNSLNIH
SEG  .....xxxxxxxxxxxxx.....
PRD  eccccccccceeeccccchhhhhhhhhhhcccccceeeccccccccceec

SEQ  YPLR
SEG  ....
PRD  cccc
```

(No Prosite data available for DKFZphtes3_2n9.2)

(No Pfam data available for DKFZphtes3_2n9.2)

DKFZphtes3_30f4

group: testes derived

DKFZphtes3_30f4 encodes a novel 192 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by LMU

Locus: /map="717.2-8 cR from top of Chr8 linkage group"

Insert length: 1388 bp

Poly A stretch at pos. 1330, polyadenylation signal at pos. 1310

```
1 CACTGAGCCC TCCTCAGATG GTTAGTGGCT TCCAACAGCC ATCAGGAGTG
51 TTTCTTGAAT GCCCCAGGTG TGGAGGACTT GGTCTGTGAC CACCTAGAAC
101 CCCAGAGCTG AACAGGAAGC CGTCCCTGCA GCAACAAGAG GGCTGGAAGG
151 GGGAGCTGCA GGCCACCCCTC GGCTCTCCCA CTGCTGGGGC GGTGATGTTT
201 GGGTGACATG TTTGAAAAAT ACTCTTAAAG ATACCAACTG TTCCCTTATA
251 TGGCTAATGG TTTGTGCAGC CACCAGCGAT GCGGGCCCTT ATTAGAGACC
301 AGGTTTGTTA AAACACCAA TATTGCTGTC CACACTAGAC ATTAACCGGC
351 TTCAGAAAAG ATGGACACCT TTTCCACGCG TGTTCGCTT CTAACTTTG
401 GTCCAGCTTT AGCCACCACA CAGCGTGTGA GGGACTGCTG CTGCGGAGTC
451 AGCCTCGTTT GTCCCTCCGC CTCCCACCAG CATGCGCCGC TTCTGAGAGA
501 CACCAGCTCC CTGCCTCCAA GCCTGGTGCC ACAGGCCTGT CGTGAGGGAC
551 CCCTGCTTCC GAGAGCTCCT GGGGGGGTTC TGCCCTTCAC CACCTGGGAG
601 AGGTGTCACT TCAGTTCCGA GTTGAACAAG GCCCGTGCAC ACAGCATGTT
651 GGGGGCCAG CCCAAAGTTC TTGTACCTC CTGATGCAAA GCCAGCCATC
701 ACCCTCCGGC CAGAGCTCAA GGTGGCCCTT TGGCCAGCCC CTCCTTGGGT
751 CCTCCAGGAG GACTGAGCAC CCCTCCTAGC GGCATCCCTT GCCCTCCACA
801 GTGCTGCCAG GGGCAGCTCG CTCTGTGCCG TGGACTGAGA CCATCCCTTG
851 GTGACAGAA TACCCCTTTG TTGAAATGC CTCGTTGCCA GAGAACTCC
901 CCAGGCATCT CGGAACGAAA CTATTTAGTT CCATTGTGAA CTGGCCACGG
951 GACAGCTTTT TATCAACTTA TTAAGTTGGA GCACTGTAAT CGCGCTTGCT
1001 GAGTTAGCAG TGGTGGTAAG CGTGTGTTAA ACACATAATG TTACGTTTAA
1051 GGAGAGAGAG GTCGTAAGGA AGTGTGCTGT CGCTCATGAC TCTCTTCTAT
1101 TAGTTGGGTA ACAGTGGCCT CATGTTTGTG TCTGTGTGTA CACAGAGCCC
1151 TTAGGTTCTG CTCTGTTTCT TTGCCAGGTG AATGTTTGTG GCATGCGCTG
1201 CTGTCCGCGC CCCTCTGTCC TGCGCAGGGT TCAGCTGTGC GCGGCCCTGA
1251 TTTCTCCAT GCACACAGAA CCTCCTTGTG TCTGTTTCTC TGTTCCTCTG
1301 TGGCTGACTC AATAAACTTT TCCCTCTGAC ATGAAAAAAA AAAAAAAA
1351 AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAG
```

BLAST Results

Entry HS549358 from database EMBL:

human STS EST67250.

Score = 2126, P = 1.5e-89, identities = 444/472

Entry HS670351 from database EMBL:

human STS WI-18501.

Score = 2089, P = 7.1e-88, identities = 445/476

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 361 bp to 936 bp; peptide length: 192

Category: putative protein

Classification: no clue


```

1 MDTFSHAVSL LNFGPALATT QRVRDCCCGV SLVCPSSHQ HAPLLRDTSS
51 LPPSLVPQAC REGPLLPRAP GGVLPTTWE RCQFSSELNK ARAHMLGAQ
101 PKVLVTSSCK ASHHPPARAQ GGPLASPSLG PPGGLSTPPS GIPCPPQCCQ
151 GHVALCRGLR PSPGDRMTRL LEMPRCQRNS PGISERNYLV PL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_30f4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_30f4, frame 1

Report for DKFZphtes3_30f4.1

```

[LENGTH]      192
[MW]           20281.56
[pI]           9.21
[BLOCKS]      BL01013C Oxysterol-binding protein family proteins
[KW]           All Alpha
[KW]           LOW_COMPLEXITY 10.94 %

SEQ  MDTFSHAVSLLNFGPALATTQRVRDCCCGVSLVCPSSHQHAPLLRDTSSSLPPSLVPQAC
SEG  .....
PRD  ccchhhhheeeccccchhhhhhccccceeeccccccccccccccccccccccccccccc

SEQ  REGPLLPRAPGGVLPTTWERQFSSELNKARAHMLGAQPKVLVTSSCKASHHPPARAQ
SEG  .....
PRD  cccccccccccccccccccccchhhhhhccccceeeccccccccccccccccccccccccc

SEQ  GGPLASPSLGPPLSTPPSGIPCPPQCCQGHVALCRGLRPSPGDRMTRLLEMPRCQRNS
SEG  xxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cccccccccccccccccccccccccchhhhhhccccceeeccccchhhccccccccccccc

SEQ  PGISERNYLVPL
SEG  .....
PRD  cccccccccccc

```

(No Prosite data available for DKFZphtes3_30f4.1)

(No Pfam data available for DKFZphtes3_30f4.1)

DKFZphtes3_35b4

group: cell cycle

DKFZphtes3_35b4 encodes a novel 1780 amino acid protein which is C-terminal identical to human M-phase phosphoprotein-1 (MPPI).

The novel protein contains a N-terminal Pfam kinesin motor domain and a ATP/GTP-binding site motif A (P-loop). MPPI is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.

The new protein can find application in modulation of the mitotic spindle.

"M-phase phosphoprotein-1" extension

motor protein

Sequenced by DKFZ

Locus: /map="750_H_1; 758_H_7; 759_C_9; 847_D_4; 906_D_1; 931_D_3; 944_C_1; 750_G_12; 800_A_11; 512.1 cR from top of Chr10 linkage group"

Insert length: 6284 bp

No poly A stretch found, no polyadenylation signal found

```
1 ATCGCAGTGC TGCTCGCGGG TCTGGCTAGT CAGGCGAAGT TTGCAGAATG
51 GAATCTAATT TTAATCAAGA GGGAGTACCT CGACCATCTT ATGTTTTTAG
101 TGCTGACCCA ATTGCAAGGC CTTGAGAAAT AAATTTTCAT GGCATTAAGC
151 TTGATCTGTC TCATGAATTT TCCTTAGTTG CTCCAATAC TGAGGCAAAAC
201 AGTTTCGAAT CTAAAGATTA TCTCCAGGTT TGTCTTCGAA TAAGACCATT
251 TACACAGTCA GAAAAAGAAC TTGAGTCTGA GGGCTGTGTG CATATTCCTGG
301 ATTCACAGAC TGTGTGCTG AAAGAGCCTC AATGCATCCT TGGTCGGTTA
351 AGTGAAAAAA GCTCAGGGCA GATGGCACAG AAATTCAGTT TTTCCAAGGT
401 TTTTGGCCCA GCAACTACAC AGAAGGAATT CTTTCAGGGT TGCATTATGC
451 AACCACTAAA AGACCTCTTG AAAGGACAGA GTCGTCTGAT TTTTACTTAC
501 GGGCTAACCA ATTCAGGAAA AACATATACA TTTCAAGGGA CAGAAGAAAA
551 TATTGCAATT CTGCCTCGAA CTTTGAATGT ATTATTGAT AGTCTTCAAG
601 AAAGACTGTA TACAAAGATG AACCTTAAAC CACATAGATC CAGAGAATAC
651 TTAAGGTTAT CATCAGAACA AGAGAAAGAA GAAATTGCTA GCAAAAGTGC
701 ATTGCTTCGG CAAATTAAG AGGTTACTGT GCATAATGAT AGTGATGATA
751 CTCTTTATGG AAGTTTAACT AACTCTTTGA ATATCTCAGA GTTTGAAGAA
801 TCCATAAAAG ATTATGAACA AGCCAACTTG AATATGGCTA ATAGTATAAA
851 ATTTTCTGTG TGGGTTTCTT TCTTTGAAAT TTACAATGAA TATATTTATG
901 ACTTATTTGT TCCTGTATCA TCTAAATTCC AAAAGAGAAA GATGCTGCGC
951 CTTTCCCAAG ACGTAAAGGG CTATTCTTTT ATAAAAGATC TACAATGGAT
1001 TCAAGTATCT GATTCCAAAG AAGCCTATAG ACTTTTAAAA CTAGGAATAA
1051 AGCACCAGAG TGTTCCTTC ACAAAATTGA ATAATGCTTC CAGTAGAAGT
1101 CACAGCATAT TCACTGTAA AATATTACAG ATTGAAGATT CTGAAATGTC
1151 TCGTGAATTT CGAGTCAGTG AATTATCTTT ATGTGATCTT GCTGGTTCAG
1201 AACGAACTAT GAAGACACAG AATGAAGGTG AAAGGTTAAG AGAGACTGGG
1251 AATATCAACA CTCTTTTATT GACTCTGGGA AAGTGTATTA ACGTCTTGAA
1301 GAATAGTGAA AAGTCAAAGT TTCAACAGCA TGTGCCTTTC CGGGAAGTA
1351 AACTGACTCA CTATTTTCAA AGTTTTTTTA ATGGTAAAGG GAAAATTTGT
1401 ATGATTGTCA ATATCAGCCA ATGTTATTTA GCCTATGATG AAACACTCAA
1451 TGTATTGAAG TTCTCCGCCA TTGCACAAAA AGTTTGTGTC CCAGACACTT
1501 TAAATTCCTC TCAAGATAAA TTATTTGGAC CTGTCAAATC TTCTCAAGAT
1551 GTATCACTAG ACAGTAATTC AAACAGTAAA ATATTAAATG TAAAAAGAGC
1601 CACCATTTCA TGGGAAAATA GTCTAGAAGA TTTGATGGAA GACGAGGATT
1651 TGGTTGAGGA GCTAGAAAAC GCTGAAGAAA CTCAAAATGT GGAAACTAAA
1701 CTTCTTGATG AAGATCTAGA TAAAACATTA GAGGAAAATA AGGCTTTTAT
1751 TAGCCACGAG GAGAAAAGAA AACTGTTGGA CTTAATAGAA GACTTGAAAA
1801 AAAAAGTATG AAATGAAAAA AAGGAAAAAT TAACCTTGGA ATTTAAAAAT
1851 CGAGAAGAAG TTACACAGGA GTTTACTCAG TATTGGGCTC AACGGGAAGC
1901 TGACTTTAAG GAGACTCTGC TTCAAGAACG AGAGATATTA GAAGAAAATG
1951 CTGAACGTCG TTTGGCTATC TTCAAGGATT TGGTTGGTAA ATGTGACACT
2001 CGAGAAGAAG CAGCGAAAAG CATTGTGTCC ACAAAGTTG AAACGGAAGA
2051 AGCTACTGCT TGTTTAGAAC TAAAGTTTAA TCAAATTTAA GCTGAATTAG
2101 CTAAGAACAA AGGAGAATTA ATCAAAACCA AAGAAGAGTT AAAAAAGAGA
2151 GAAAATGAAT CAGATTCATT GATTCAAGAG CTTGAGACAT CTAATAAGAA
2201 AATAAATTACA CAGAATCAAA GAATTAAGA ATTGATAAAT ATAATTGATC
2251 AAAAAGAAAG TACTATCAAC GAATTCAGA ACCTAAAGTC TCATATGGAA
2301 AACACATTTA AATGCAATGA CAAGGCTGAT ACATCTTCTT TAATAATAAA
2351 CAATAAATTG ATTTGTAATG AAACAGTTGA AGTACCTAAG GACAGCAAAT
2401 CTAAGATCTG TTCAAGAAAG AAAAGAGTAA ATGAAAATGA ACTTCAGCAA
2451 GATGAACCCG CAGCAAAGAA AGGGTCTATC CATGTTAGTT CAGTATCAC
2501 TGAAGACCAA AAGAAAAGTG AAGAAGTGGC ACCGAACATT CGAGAAATTG
2551 AAGACATCAG AGTTTACAAA GAAAATAATG AAGGACTGAG AGCATTTTTA
```



```

2601 CTCACTATTG AGAATGAACT TAAAAATGAA AAGGAAGAAA AAGCAGAATT
2651 AAATAAACAG ATTGTTCAAT TTCAGCAGGA ACTTTCTCTT TCTGAAAAAA
2701 AGAATTTAAC TTAAAGTAAA GAGGTCCAAC AAATTCAGTC AAATTATGAT
2751 ATTGCAATTG CTGAATTACA TGTGCAGAAA AGTAAAAATC AAGAACAGGA
2801 GGAAGAGATC ATGAAATTGT CAAATGAGAT AGAAACTGCT ACAAGAAGCA
2851 TTACAAATAA TGTTCACAA ATAAATTTAA TGCACACGAA AATAGACGAA
2901 CTACGTACTC TTGATTCAGT TTCTCAGATT TCAACATAG ATTGTCTCAA
2951 TCTCAGGGAT CTGTCAAATG GTTCTGAGGA GGATAATTG CCAAATACAC
3001 AGTTAGACCT TTTAGGTAAT GATTATTTGG TAAGTAAGCA AGTTAAAGAA
3051 TATCGAATTC AAGAACCCTA TAGGGAAAAT TCTTTCCACT CTAGTATTGA
3101 AGCTATTTGG GAAGAATGTA AAGAGATTGT GAAGGCCTCT TCCAAAAAAA
3151 GTCATCAGAT TGAGGAAGCT GAACAACAAA TTGAAAAATT GCAGGCAGAA
3201 GTAAAGGCTT ATAAGGATGA AAACATAGA CTAAAGGAGA AGGAGCATAA
3251 AAACCAAGAT GACCTACTAA AAGAAAAAGA AACTCTTATA CAGCAGCTGA
3301 AAGAAGAATT GCAAGAAAAA AATGTTACTC TTGATGTTCA AATACAGCAT
3351 GTAGTTGAAG GAAAGAGAGC GCTTTCAGAA CTACACAAAG GTGTTACTTG
3401 CTATAAGGCA AAAATAAAGG AACTTGAAAC AATTTTAGAG ACTCAGAAAG
3451 TTGAACGTAG TCATTACGCC AAGTTAGAAC AAGACATTTT GGAAGAGGAA
3501 TCTATCATCT TAAAGCTAGA AAGAAATTTG AAGGAATTTT AAGAACATCT
3551 TCAGGATTTCT GTCAAAAACA CCAAGATTTT AAATGTAAAG GAACTCAAGC
3601 TGAAGAAGA AATCACACAG TTAACAAATA ATTTGCAAGA TATGAAACAT
3651 TTACTTCAAT TAAAAGAAGA AGAAGAAGAA ACCAACAGGC AAGAACAGA
3701 AAAATTTGAA GAGGAAGCTC CTGCAAGCTC TGCTCGTACC CAGAATCTGA
3751 AAGCAGATCT TCAGAGGAAG GAAGAAGATT ATGCTGACCT GAAAGAGAAA
3801 CTGACTGATG CCAAAAAGCA GATTAAGCAA GTACAGAAAG AGGTATCTGT
3851 AATGCGTGAT GAGGATAAAT TACTGAGGAT TAAAAATTAAT GAACTGGAGA
3901 AAAAGAAAAA CCAGTGTCTC CAGGAATTAG ATATGAAGCA GCGAACCAT
3951 CAGCAACTCA AGGAGCAGTT AAATAATCAG AAAGTGGAAAG AAGCTATACA
4001 ACAGTATGAG AGAGCATGCA AAGATCTAAA TGTAAAGAG AAAATAATTG
4051 AAGACATGGC AATGACACTA GAAGAACAGG AACAACTCA GGTAGAACAG
4101 GATCAAGTGC TTGAGGCTAA ATTAGAGGAA GTTGAAAGGC TGGCCACAGA
4151 ATTGGAAAAA TGAAGGAAA AATGCAATGA TTTGGAACCC AAAAACCAATC
4201 AAAGGTCAAA TAAAGAACAT GAGAACAAAC CAGATGTGCT TGGAAAGCTC
4251 ACTAATCTTC AAGATGAGTT ACAGGAGTCT GAACAGAAAT ATAATGCTGA
4301 TAGAAAGAAA TGGTTAGAAG AAAAAATGAT GCTTATCACT CAAGCGAAAG
4351 AAGCAGAGAA TATACGAAAT AAAGAGATGA AAAAAATATG TGAGGACAGG
4401 GAGCGTTTTT TTAAGCAACA GAATGAAATG GAAATACTGA CAGCCCAGCT
4451 GACAGAGAAA GATAGTGACC TTCAAAGAGT GCGAGAAGAA CGAGATCAAC
4501 TGGTTGCAGC TTTAGAAATA CAGCTAAAAG CACTGATATC CAGTAATGTA
4551 CAGAAGATA ATGAAATTGA ACAACTAAAA AGGATCATAT CAGAGACTTC
4601 TAAATAGAAA ACACAAATCA TGGATATCAA GCCCAAACGT ATTAGTTTCA
4651 CAGATCCTGA CAAACTTCAA ACTGAACCTC TATCGACAAG TTTTGAATTT
4701 TCCAGAAATA AAATAGAGGA TGGATCTGTA GTCCCTGACT CTTGTGAAGT
4751 GTCAACAGAA AATGATCAAA GCACTCGATT TCCAAAACCT GAGTTAGAGA
4801 TTCAATCTTC ACCTTTACAG CCAACAAAAA TGGCAGTGAA ACACCCCTGGT
4851 TGTACCACAC CAGTGACAGT TGAGATTCCC AAGGCTCGGA AGAGGAAGAG
4901 TAATGAAATG GAGGAGGACT TGGTGAAATG TGAAAATAAG AAGAATGCTA
4951 CACCCAGAAC TAAATTTGAAA TTTCTTATTT CAGATGATAG AAATTTCTCT
5001 GTCAAAAAGG AACAAAAGGT TGCCATACGT CCATCATCTA AGAAAACATA
5051 TTCTTTACGG AGTCAGGCAT CCATAATTGG TGTAAACCTG GCCCTAAGA
5101 AAAAAAGAAG AACACTACAG AAATTTGGAG ACTTCTTACA ACATTCTCCC
5151 TCAATTTCTC AATCAAAAGC AAAGAAGATA ATTGAAACAA TGAGCTCTTC
5201 AAAGCTCTCA AATGTAGAAG CAAGTAAAGA AAATGTGTCT CAACCAAAAC
5251 GAGCCAAACG GAAATTATAC ACAAGTGAAA TTTCATCTCC TATTGATATA
5301 TCAGGCCAAG TGATTTTAAT GGACCAGAAA ATGAAGGAGA GTGATCACCA
5351 GATTATCAAA CGACGACTTC GAACAAAAAC AGCCAAATAA ATCACTTATG
5401 GAAATGTTTA ATATAAATTT TATAGTCATA GTCATTGGAA CTTGCATCCT
5451 GTATTGTAAA TATAAATGTA TATATTATGC ATTAAATCAC TCTGCATATA
5501 GATTGCTGTT TTATACATAG TATAATTTTA ATTCAATAAA TGAGTCAAAA
5551 TTTGTATATT TTTATAAGGC TTTTATATA TAGCTTCTTT CAAACTGTAT
5601 TTCCCTATTA TCTCAGACAT TGGATCAGTG AAGATCCTAG GAAAGAGGCT
5651 GTTATTCTCA TTTATTTTGC TATACAGGAT GTAATAGGTC AGGTATTTGG
5701 TTTACTTATA TTTAACAATG TCTTATGAAT TTTTCTTACT TTATCTGTTA
5751 TACAACGTAT TTTACATATC TGTTTGGATT ATAGCTAGGA TTTGGAGAA
5801 AAGTGTGTAC AGATCAGAAA ACATGTATAT ACATTATTTA GAAAAGATCT
5851 CAAGTCTTTA ATTAGAATGT CTCACTTATT TTGTAACAT TTTGTGGGTA
5901 CATAGTACAT GTATATATTT ACGGGGTATG TGAGATGTTT TGACACAGGC
5951 ATGCAATGTG AAATACGTGT ATCATGGAGA ATGAGGTATC CATCCCCTCA
6001 AGCATTTTTC CTTTGAATTA CAGATAATCC AATTACATTC TTTAGATCAT
6051 TTAATAATAT ACAAGTAAGT TATTATTGAT TATAGTCACT CTATTGTGCT
6101 ATCAGATAGT AGATCATTCT TTTTATCTTA TTTGTTTTTG TACCCATTAA
6151 CCATCCCCAC CTCCCCCTGC AACCCTCAGT ACCCTTACCA GCCACTGGTA
6201 ACCATTCTTC TACTCTGTAT GCCCATGAGG TCAATTGATT TTATTTTATG
6251 ATCCCATAAA TAAATGAGAA CATGCAAAAA AAAA

```

BLAST Results

Entry HS898149 from database EMBL:
human STS WI-9217.

Score = 4247, P = 1.5e-187, identities = 855/862

Medline entries

94119956:

Cloning of cDNAs for M-phase phosphoproteins recognized by the MPM2 monoclonal antibody and determination of the phosphorylated epitope.

98101856:

Interaction of a Golgi-associated kinesin-like protein with Rab6.

95122643:

Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis.

Peptide information for frame 3

ORF from 48 bp to 5387 bp; peptide length: 1780

Category: known protein

Classification: Cell structure/motility

Prosites motifs: ATP_GTP_A (152-160)

```

1 MESNFNQEGV PRPSYVFSAD PIARPSEINF DGIKLDLSHE FSLVAPNTEA
51 NSFESKDYLO VCLRIRPETQ SEKELESEGC VHILDSQTVV LKEPQCILGR
101 LSEKSSGQMA QKFSFSKVFG PATTQKEFFQ GCIMQPVKDL LKGQSRIFT
151 YGLTNSGKTY TFQGTENIG ILPRTLNVLF DSLQERLYTK MNLKPHRSRE
201 YLRISSEQEK EEIASKSALL ROIKEVTVHN DSDDTLYGSL TNSLNISEFE
251 ESIKDYEQAN LNMANSIKFS VVVSFFEIYN EYIYDLFVPV SSKFQKRKML
301 RLSQOVKGYS FIKDLQWIVQ SDSKEAYRLL KLGIKHQSVA FTKLNNASSR
351 SHSIFTVKIL QIEDSEMSRV IRVSELSLCD LAGSERTMKT QNEGERLRET
401 GNINTSLTLT GKCNVULKNS ESKFQQHVP FRESKLTHYF QSFFNGKGKI
451 CMIVNISQCY LAYDETLNVL KFSIAQKVC VPDTLNSSQD KLFQPVKSSQ
501 DVSLDSNSNS KILNVKRATI SWENSLEDLM EDEDLVEELE NAEETQNVET
551 KLLEDLDLKT LEENKAFISH EEKRKLLDLI EDLKKKLINE KKEKLTLEFK
601 IREEVTQEFY QYWAQREADF KETLLQEREI LEENAERRLA IFKDLVGKCD
651 TREEAARDIC ATKVETEEAT ACLELKFNQI KAEAKTKGE LIKTKEELKK
701 RENESDSLII ELETSNKKII TQNQRIKELI NIIDQKEDTI NEFQNLKSHM
751 ENTFCNCOKA DTSSLIINNK LICNETVEVP KDSKSKICSE RKRNVNENELQ
801 QDEPPAKKGS IHVSSAIED QKKSEEVPRN IAEIEDIRVL QENNEGLRAF
851 LLTIENELKN EKEERAELNK QIVHFQOELS LSEKKNLTLS KEVQQIQSNY
901 DIAIAELHWQ KSKNQEQEEK IMKLSNEIET ATRSITNNVS QIKLMHTKID
951 ELRTLDVSQS ISNIDLLNLR DLSNGSEEDN LPNTQLDOLL NDYLVSKQVK
1001 EYRIQEPNRE NSFHSSIEAI WEECKEIVKA SSKKSHQIEE LEQQIEKLQA
1051 EVKGYKDENN RLKEKEHKNO DDLLEKETL IQQLKEELQE KNVTLDVQIQ
1101 HVVEGKRALS ELTQGVTCYK AKIKELETIL ETQKVERSHS AKLEQDILEK
1151 ESIIKLKERN LKEFQEHLOD SVKNTKDLNV KELKLKEEIT QLTNNLQDMK
1201 HLLQLKEEEE ETNRQETEKL KEELSASSAR TQNLKADLQR KEEDYADLKE
1251 KLTDAAKQIK QVQKEVSVMR DEDKLLRIKI NELEKKNQOC SQELDMKQRT
1301 IQQLKEQLNN QKVEEAIQY ERACKDLNVK EKIIEDMRMT LEEQEQTQVE
1351 QDOVLEAKLE EVERLATELE KWKEKCNLE TKNNQRSNKE HENNTDVLGK
1401 LTNLQDELQE SEQYNADRK KWLEEKMMLI TQAKEAENIR NKEMKKYAED
1451 RERFFKQONE MEILTAQLTE KSDQLQWRE ERDQLVALE IQLKALISSN
1501 VQKDNEIEQL KRISSETSKI ETQIMDIKPK RISSADPKL QTEPLSTSFE
1551 ISRNKIEDGS VVLDSECVST ENDQSTRFPK PELEIQFTPL QPNKMAVKHP
1601 GCTTPVTVEI PKARKRSNE MEEDLVKCN KKNATPRTNL KFPISDDRNS
1651 SVKKEQKVAI RPSSKKTYSL RSQASIIQVN LATKKKEGTL QKFGDFLQHS
1701 PSILQSKAKK IETMSSSKL SNVEASKENV SQPKRAKRKL YTSEISSPID
1751 ISGQVILMDQ KMKESDHQII KRRLRTKTAK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_35b4, frame 3

TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds., N = 1, Score = 3743, P = 0

PIR:A36881 MPM2-reactive phosphoprotein 1 - human (fragment), N = 2,
Score = 2808, P = 2.5e-294

TREMBL:AF070672_1 product: "rabkinesin6"; Homo sapiens rabkinesin6
mRNA, complete cds., N = 2, Score = 680, P = 2.6e-99

>TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase
phosphoprotein-1 mRNA, partial cds.
Length = 753

HSPs:

Score = 3743 (561.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 752/753 (99%), Positives = 753/753 (100%)

Query: 1028 VKASSKKSHQIEELEQQIEKLAQEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE 1087
VKASSKKSHQIEELEQQIEKLAQEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE
Sbjct: 1 VKASSKKSHQIEELEQQIEKLAQEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE 60

Query: 1088 LQEKNVTLDVQIQHVVEGKRALSELQTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
LQEKNVTLDVQIQHVVEGKRALSELQTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI
Sbjct: 61 LQEKNVTLDVQIQHVVEGKRALSELQTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 120

Query: 1148 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 1207
LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE
Sbjct: 121 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 180

Query: 1208 EEEETNRQETEKLEELSASSARTQNLKADLQKKEEDYADLKEKLTDAKKQIKQVQKEVS 1267
EEEETNRQETEKLEELSASSARTQNLKADLQKKEEDYADLKEKLTDAKKQIKQVQKEVS
Sbjct: 181 EEEETNRQETEKLEELSASSARTQNLKADLQKKEEDYADLKEKLTDAKKQIKQVQKEVS 240

Query: 1268 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL 1327
VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL
Sbjct: 241 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL 300

Query: 1328 NVKEKIIEDMRMTLEEEOEQTQVEQDQVLEAKLEEVEERLATELEKWKKEKNDLETKNQRS 1387
NVKEKIIEDMRMTLEEEOEQTQVEQDQVLEAKLEEVEERLATELEKWKKEKNDLETKNQRS
Sbjct: 301 NVKEKIIEDMRMTLEEEOEQTQVEQDQVLEAKLEEVEERLATELEKWKKEKNDLETKNQRS 360

Query: 1388 NKEHENNTDVLGKLTNLQDELQSESEQYNADRKKWLEEKMLITQAKEAENIRNKEMKKY 1447
NKEHENNTDVLGKLTNLQDELQSESEQYNADRKKWLEEKMLITQAKEAENIRNKEMKKY
Sbjct: 361 NKEHENNTDVLGKLTNLQDELQSESEQYNADRKKWLEEKMLITQAKEAENIRNKEMKKY 420

Query: 1448 AEDRERFFKQQNEMEILTAQLTEKDSDLQKWEERDQLVAALIEIQLKALISSNVQKDNEI 1507
AEDRERFFKQQNEMEILTAQLTEKDSDLQKWEERDQLVAALIEIQLKALISSNVQKDNEI
Sbjct: 421 AEDRERFFKQQNEMEILTAQLTEKDSDLQKWEERDQLVAALIEIQLKALISSNVQKDNEI 480

Query: 1508 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGSVVLDSCE 1567
EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGSVVLDSCE
Sbjct: 481 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGSVVLDSCE 540

Query: 1568 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKS NEMEEDLVK 1627
VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTV+IPKARKRKS NEMEEDLVK
Sbjct: 541 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVKIPKARKRKS NEMEEDLVK 600

Query: 1628 CENKKNATPRTNLKFPISDDRNSVVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE 1687
CENKKNATPRTNLKFPISDDRNSVVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE
Sbjct: 601 CENKKNATPRTNLKFPISDDRNSVVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE 660

Query: 1688 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 1747
GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS
Sbjct: 661 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 720

Query: 1748 PIDISGQVILMDQMKKESDHQIIKRRRLRTKTAK 1780
PIDISGQVILMDQMKKESDHQIIKRRRLRTKTAK
Sbjct: 721 PIDISGQVILMDQMKKESDHQIIKRRRLRTKTAK 753

Score = 197 (29.6 bits), Expect = 2.1e-11, P = 2.1e-11
Identities = 114/542 (21%), Positives = 253/542 (46%)

Query: 692 IKTKELKKRENESDSLQLETSNKKIITQNRKELINIIDQKEDTINEFONLKSMM- 750
+K + + E + I++L+ K +N R+KE + ++D + E + L +
Sbjct: 1 VKASSKKSHQIEELEQQIEKLAQEVKGYKDENNRLKEKEH--KNQDOLLKEKETLIQQLK 58

Query: 751 ENTFKCNDAKADTS-SLIINNKLICNETVEVPKDSKSKICSERKRVNENELQQDEPPAK-- 807
E + N D ++ K +E + K+KI E + + E + + AK
Sbjct: 59 EELQEKNVTLDVQIQHVVEGKRALSELQTQGVTCYKAKI-KELETILETQKVERSHSAKLE 117

Query: 808 KGSIHVSSAITEDQKKSEEVPRNIAE-IEDIRVLQENNEGLRAFLLLTIENELKNEK---- 862

+ + S I + ++ +E + ++ + +++ + L L+ + + N L++ K
 Sbjct: 118 QDILEKESIIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQ 177
 Query: 863 --EEKAELNQIVH-FQELSLSEKKNLTLSKEVQIQSNYDIAIAELHVQSKNQEQEE 919
 EE+ E N+Q ++ELS S + L ++Q+ + Y A+L K K + ++
 Sbjct: 178 LKEEEETNRQETEKLEELSASSARTQNLKADLQRKEEDY----ADL---KEKLTDAKK 230
 Query: 920 KIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDVSQISNIOLLNLRDLSNGSEE 978
 +I ++ E+ S+ + + KL+ KI+EL + + SQ +D+ R + E+
 Sbjct: 231 QIKQVQKEV-----SVMRD--EDKLLRIKINELEKKNQCSQ--ELDMKQ-RTIQQLKEQ 280
 Query: 979 DNLPTQLDOLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAIWEECKEIVKASSKKSHQI 1038
 N N +++ Y + K+ ++E E+ ++E + E + K ++
 Sbjct: 281 LN--NQKVEEAIQY--ERACKDLNVKEKIIED-MRMTLEEQTQVEQDQVLEAKLEEV 335
 Query: 1039 EELEQIEKLQAEVKGKIDENNRLEKEKHNQDDLKEKETLIQQLKEELQEKNTV---- 1094
 E L ++EK + + + +NN+ KEH+N D+L + L +L+E Q+ N
 Sbjct: 336 ERLATELEKWKECNDLETNNQRSNKEHENNTDVLGKLTNLQDELQESQKYNADRRKKW 395
 Query: 1095 LDVQIQHVVEGKRA-----LSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
 L+ ++ + + K A + + + + + E+E IL Q E+ + ++
 Sbjct: 396 LEEKMLITQAKEAENIRNKEMKKAEDRERFFKQONEME-ILTAQLTEKSDQKQWRE- 453
 Query: 1148 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELK-LKEEITQLTNNLQDMKHLQLK 1206
 E++ ++ LE LK + +V+ KD +++LK + E +++ + D+K +
 Sbjct: 454 -ERDQLVAALQKAL---ISSNVQ--KDNEIEQLKRIISETSKIETQIMDIK---PKR 504
 Query: 1207 EEEETNRQETEKLEELSASSARTQN 1233
 + ++ +TE L S + ++
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIED 531
 Score = 186 (27.9 bits), Expect = 3.2e-10, P = 3.2e-10
 Identities = 131/674 (19%), Positives = 294/674 (43%)
 Query: 673 LELKFNQIKAEAKTKGELIKT-KEELKKRENESDSLIQELETSSNKKIITQNRIKELIN 731
 L+ K ++ + +L K K LI+ KEEL+++ D IQ + + + Q +
 Sbjct: 35 LKEKEKHNQDDLKEKETLIQQLKEELQEKNTLDVQIQHVVEGKRALSELTOGVTCYKA 94
 Query: 732 IIDQKEDTINEFQNL-KSHMENTFKCNDKADTSSLIINNKLICNETVEVPKDSKSKICSE 790
 I + E TI E Q + +SH + D + S+I+ + E E +DS
 Sbjct: 95 KIKELE-TILETQKVERSHSAKLEQ--DILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKKEEETNRQETEKLEELSASSARTQNL 207
 Query: 791 RKRNVENELQ-QDEPPAKKGSIHVSSAITEDQKKEEV-RPNIAEI-EDIRVLQENNEGL 847
 K +N EL+ ++E ++ + + +++ EE R ++ E++ + L
 Sbjct: 148 TKDLNVKELKLKEEITQLTNNLQDMKHLQLKKEEETNRQETEKLEELSASSARTQNL 207
 Query: 848 RAFLTIENELKNEKEEKAELNQIVHFQELSLSEKKNLTLSKEVQIQI-----QSNYDI 902
 +A L E + + KE+ + KQI Q+E+S+ ++ L ++ ++ Q + ++
 Sbjct: 208 KADLQRKEEDYADLKEKLTDAKKQIKQVQKEVSVMRDEKLLRIKINELEKKNQCSQEL 267
 Query: 903 AIAELHVQSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDVSQI 961
 + + +Q+ K Q +K+ + + E A + + I+ M ++E +T Q+
 Sbjct: 268 DMKQRTIQQLKEQLNNQKVEEAIQYERACKDLNVKEKIIEDMRMTLEEQTQVEQDQV 327
 Query: 962 SNIDLLNLRDLSNGSEEDNLPNTQLDOLLGNDYLVSKQVKEYRI--QEPNRENSFHSSIEA 1019
 L + L+ E+ L+ N + + + N ++ S +
 Sbjct: 328 LEAKLEEVLATELEKWKECNDLETNNQRSNKEHENNTDVLGKLTNLQDELQESQK 387
 Query: 1020 IWEECKEIVKASSKKSHQIEELEQIEKLQAEVKGKIDENNRLEKEKHNQ--DDLKEK 1077
 + K+ ++ Q +E E K E+K Y ++ R +++++ + L EK
 Sbjct: 388 YNADRRKKWLEEKMLITQAKEAENIRNK---EMKKAEDRERFFKQONEMEILTAQLTEK 444
 Query: 1078 ETLIQQLKEELQEKNTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKVER 1137
 ++ +Q+ +EE + L++Q++ ++ + + ++ +ET + K +R
 Sbjct: 445 DSDQLQWREERDQLVAALQKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKR 504
 Query: 1138 SHSAKLEQDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLT 1193
 SA ++ E S ++ RN E + DS +N + + +L+ + T L
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIEDGVSVDSCVSTENDQSTRFPKPELEIQTPLQ 564
 Query: 1194 NNLQDMKH---LLQLKEEETNRQETEKLEEL-SASSARTQNLKADLQRKEEDYADLK 1249
 N +KH + + + +++ +E+L + + + +L+ D +
 Sbjct: 565 PNKMAVKHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKNATPRTNLKFPISDDRNSS 624
 Query: 1250 EKLTDKQIKQVQKEVSVMRDEKLLRIKINELEKKNQCSQEL-DMKQRTIQQLKEQL 1308
 K + K I+ K+ +R+ + I +N KKK Q+ D Q + L+ +
 Sbjct: 625 VK-KEQKVAIRPSSKKTYSLSQASI--IGVNLATKKKEGTLQKFGDFLOHSPSILQSKA 681
 Query: 1309 NNQKVEEAIQYERACKDLNVKEKIIEDMR 1338
 +K+ E + + + + + KE + + R
 Sbjct: 682 --KKIETMSSSKLSNVEAS-KENVSQPKR 708

Score = 165 (24.8 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 140/626 (22%), Positives = 271/626 (43%)

Query: 536 VEELENAEETQNVETKLLDEDLDKLTLEENKAFISHEEKRLLDLIEDLKKKLINEKKEK- 594
+EELE E E K +D+ L+E+ H+ + LL E L ++L E +EK
Sbjct: 11 IEELEQQIEKLQAEVKGY-KDENNRLEKE-----HKNQDDLLKEKETLIQQLKEELQEK 65

Query: 595 LTLEFKIREEVT-----QEFTQYWAQREADFKE--TLQEREILEENAERRLAIFKDLVG 647
+TL+ +I+ V E TQ +A KE T+L+ +++ E + +L +D++
Sbjct: 66 VTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV-ERSHSKLE--QDILE 122

Query: 648 KCDT---REEAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENE 704
K E K+ ++ + T L +K ++K E+ + L K L+ +E E
Sbjct: 123 KESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKEE 182

Query: 705 SDSLIQELETSSNKKIITQNRKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSS 764
++ QE E +++ + R + L + +KE+ + + + + K K + S
Sbjct: 183 EETNRQETEKLEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQK-EVSV 241

Query: 765 LIINNKLICNETVEVPKDSKSKICSERKRVNENELQODEPPAKKGSIHVSSAITEQKKS 824
+ +KL+ + E+ K K CS+ + + +QQ+ + V AI + + +
Sbjct: 242 MRDEDKLLRIKINELEK--KKNQCSQELDMKQRTIQQLKEQLNNQK--VEEAIQYERAC 297

Query: 825 EEVPRNIAEIEDIRVLQENNEGLRAFLITENELKNEKEEKAELNKQIVHFQQLSLSEK 884
+++ IED+R+ E E + + + L+ + EE L ++ +++ + E
Sbjct: 298 KDLNVKEKIIDMRMTLEEQEQTQ---VEQDQVLEAKLEEVEERLATELEKWKCKNDLET 354

Query: 885 KNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNE-IETATRSITN-----N 938
KN S + + ++N D+ + +L + + QE E+K + +E IT N
Sbjct: 355 KNNQRSNK--EHENNTDV-LGKLTNLQDELQSEQKYNADRRKKWLEKMMILITQAKEAEN 411

Query: 939 VSQIKLMHTKIDELRTLDVSQISNIDL-LNLRD--LSNGSEEDNLPNTQLDLLGNDYLV 995
+ ++ D R +++ + L +D L EE + L++ +
Sbjct: 412 IRNKEMKKYAEADRERFFKQONEMEILTAQLTEKSDQLQWREERDQLVALEIQLKALIS 471

Query: 996 SKQVKEYRIQEPNRENSFHSSIEA-IWE-ECKEIVKASSKKSHQIEELEQQIEKLQAEVK 1053
S K+ I++ R S S IE I + + K I A K Q E L E + +++
Sbjct: 472 SNVQKDNIEQLKRIISETSKIETQIMDIKPKRISSADPKL-QTEPLSTSFEISRKNIE 530

Query: 1054 GYKDNNRLKEKEHKNQDDLLKEKE-----TLIQLKEELQEKNVTLDVQIQHVVEGKRA 1108
+ + + +Q + E T +Q K ++ T V ++ KR
Sbjct: 531 DGSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVHPGCTTPVTVKIPKARKRK 590

Query: 1109 LSELTQG-VTCYKAKIKELETILETQ-KVERSHSAKLEQDILEKES 1152
+E+ + V C K T L+ +R+ S K EQ + + S
Sbjct: 591 SNEMEEDLVKCNKKNATPRTNLKFPISDDRNSSVKKQKQVAIRPS 636

Score = 143 (21.5 bits), Expect = 1.3e-05, P = 1.3e-05
Identities = 164/684 (23%), Positives = 304/684 (44%)

Query: 295 QKRKMLR-LSQDVKGYSFIKDLQWQVSDSKEAYRLKLGIKHQSVAFTKLNNASS---- 349
+K +++ L +++ + D+Q V + K A L G+ +L
Sbjct: 49 EKETLIQLKEELQEKNVTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV 108

Query: 350 -RSHSI-FTVKILQIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGE-RLRETGNINTS 406
RSHS IL+ E + + E L S + K N E +L+E T+
Sbjct: 109 ERSKSAKLEQDILEKESIILKLERNLKEFQE-HLQDSVKNTKDLNVKELKLKEEITQLTN 167

Query: 407 LLTLGKCINVLNSEKSKFQHVFPRESKLTHYFQSFFNGKGRKICMIVNISQCYLAYDET 466
L K + LK E+ +Q + +L+ N K + + Y E
Sbjct: 168 NLQDMKHLQLKEEEETNRQETEKLEELSASSARTQNLKADL---QRKEEDYADLKEK 224

Query: 467 LNVLFKSAIAQKVCVPTDNLSSQDKLFGPVKSSQDVSLDSNSKILNVKRATISWENSL 526
L K I Q V ++ +DKL +K ++ + N S+ L++K+ TI
Sbjct: 225 LTDAK-KQIKQ-VQKEVSVMRDEKLLR-IKINE-LEKKNQCSQELDMKQRTIQQLKEQ 280

Query: 527 EDLMEDEDLVEELENAEETQNVETKLLDEDLDKLTLEENKAFISHEEKRLLDL-IEDLKK 585
+ + E+ +++ E A + NV+ K++ ED+ TLEE + + E+ ++L+ +E++++
Sbjct: 281 LNNQKVEEAIQYERACKDLNVKEKII-EDMRMTLEEQE--TQVEQDQVLEAKLEEVEER 337

Query: 586 KLIN-EK-KEKLT-LEFKIREEVTQEFTQYWAQREADFKETLLQEREILEE-----NAERR 638
EK KEK LE K + +E + K T LQ+ E+ E NA+R+
Sbjct: 338 LATELEKWKCKNDLETNNQRSNKEHEN---NTDVLGKLTNLQD-ELQSEQKYNADRK 393

Query: 639 LAIFKDLVGKCDTREAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEEL 698
+ + ++ T+ + A++I KE ++ E F Q + E+ +L + +L
Sbjct: 394 KWLEEKMM--LITQAKEAENI-RNK-EMKKYAEADRERFFKQ-QNEMEILTAQLTEKSDSL 448

Query: 699 KKRENESDSLIQELETSSNKKIITQN-QR---IKELINIIDQKEDTINEFQNLKSHMENTF 754
+K E D L+ LE K +I+ N Q+ I++L II + + ++K ++

Sbjct: 449 QKWREERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSA 508

Query: 755 KCNDKADTSSLIINNKLICN--ETVEVPKDSKSKICSERK---RVNENELQ-QDEP--PA 806
DK T L + ++ N E V DS ++ +E R + EL+ Q P P

Sbjct: 509 D-PDKLTEPLSTSFEISRNKIEDGSVVLDS-CEVSTENDQSTREFPKPELEIQFTPLQPN 566

Query: 807 KKGSIH--VSSAITEDQKKSEEVPRNIAEIEDIRVLQENNEGLRA---FLTTIENELKNE 861
K H ++ +T K+ ++ N E + ++ + N R F ++ + +

Sbjct: 567 KMAVKHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKKNATPRTNLKFPISDDRNSVK 626

Query: 862 KEEKAEL---NKQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQE 918
KE+K + +K+ + + S+ NL K+ +Q D + +SK ++

Sbjct: 627 KEQKVAIRPSSKKTYSLSRSQASIIGV-NLATKKKEGTLQKFGDFLQHSFSLQSKAKKII 685

Query: 919 EKIM--KLSNEIETATRSITNNVSQIKLMHTKI--DELRT-LDSVSQISNID 965
E + KLSN +E + NVSQ K K+ E+ + +D Q+ +D

Sbjct: 686 ETMSSSKLSN-VEASKE----NVSQPKRAKRKLYTSEISSPIDISGQVILMD 732

Score = 133 (20.0 bits), Expect = 1.6e-04, P = 1.6e-04
Identities = 94/426 (22%), Positives = 188/426 (44%)

Query: 527 EDLM-EDEDLVEELENAEETQNVETKLLDDELDKTLEENKAFISHEERKLLDL-IEDLK 584
+DL+ E E L+++L+ + +NV LD + +E +A + I++L+

Sbjct: 44 DLLLKEKETLIQQLKEELQEKQNT---LDVQIQHVVEGKRALSELTOGVTCTYKAKIKELE 100

Query: 585 KKLINEKKEKLTLEFKIREEVQ-EFTQYWAQREA-DFKETLLQEREILEENAERRLAIF 642
L +K E+ + K+ +++ + E +R +F+E L + ++ + L +

Sbjct: 101 TILETQKVER-SHSAKLEQDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKL- 158

Query: 643 KDLVGKCDTREAAKDICATKVETEATACLELKFNQIKAEAKTKGELIKTKEELKKRE 702
K+ + + + K + K E EE + ++ EL+ + K +L+++E

Sbjct: 159 KEEITQLTNNLQDMKHLQLKEEEETN---RQETEKLEELSASSARTQNLKADLQKKE 215

Query: 703 NESDSLQLETSNKKIITQNRKIKELINIIDQK-EDTINEFQNLKSHMENTFKCNDKA- 760
+ L ++L T KK I Q Q+ ++ D+ INE + K+ +

Sbjct: 216 EDYADLKEKL-TDAKKQIKQVQKEVSVMRDEDKLLRIKINELEKKKNQCSQELDMKQRTI 274

Query: 761 DTSSLIINNKLICNETVE---VPKDS--RKSICSE-RKRVNENE---LQDEPPAKKGS 810
+NN+ + E ++ KD K KI + R + E E ++QD+ K

Sbjct: 275 QQLKEQLNNQKV-EBAIQYERACKDLNVKEKIEDMRMTLEEQQEQTQVEQDQVLEAKLE 333

Query: 811 IHVSSAITEDQKKSEEVPR-NIAEIEDIRVLQENNEGLRAFLLTTIENELKNEKEEKAELN 869
V TE +K E+ + ENN + L +++EL+ E E+K +

Sbjct: 334 -EVERLATELEKWKCKCNDLETNNQSRNKEHENNTDVLGKLTNLQDELQ-ESEQKYNAD 391

Query: 870 KQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNEIE 929
++ ++++ L +T +KE + I++ + K E E+ K NE+E

Sbjct: 392 RK-KWLEEKMMML-----ITQAKEAENIRNK-----EMKKAEDRERFFKQQNEME 435

Query: 930 TATRSITNNVSQIKLMHTKIDEL 952
T +T S ++ + D+L

Sbjct: 436 ILTAQLTEKDSDLQKWREERDQL 458

Pedant information for DKFZphtes3_35b4, frame 3

Report for DKFZphtes3_35b4.3

[LENGTH] 1780
[MW] 206176.77
[pI] 5.60
[HOMOL] TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds. 0.0
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) (S. cerevisiae, YDL058w) 7e-30
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-21